



*Enabling today.  
Inspiring tomorrow.*

# ENERGY EFFICIENT GRAPHICS AND MULTIMEDIA IN 28NM CARRIZO APU

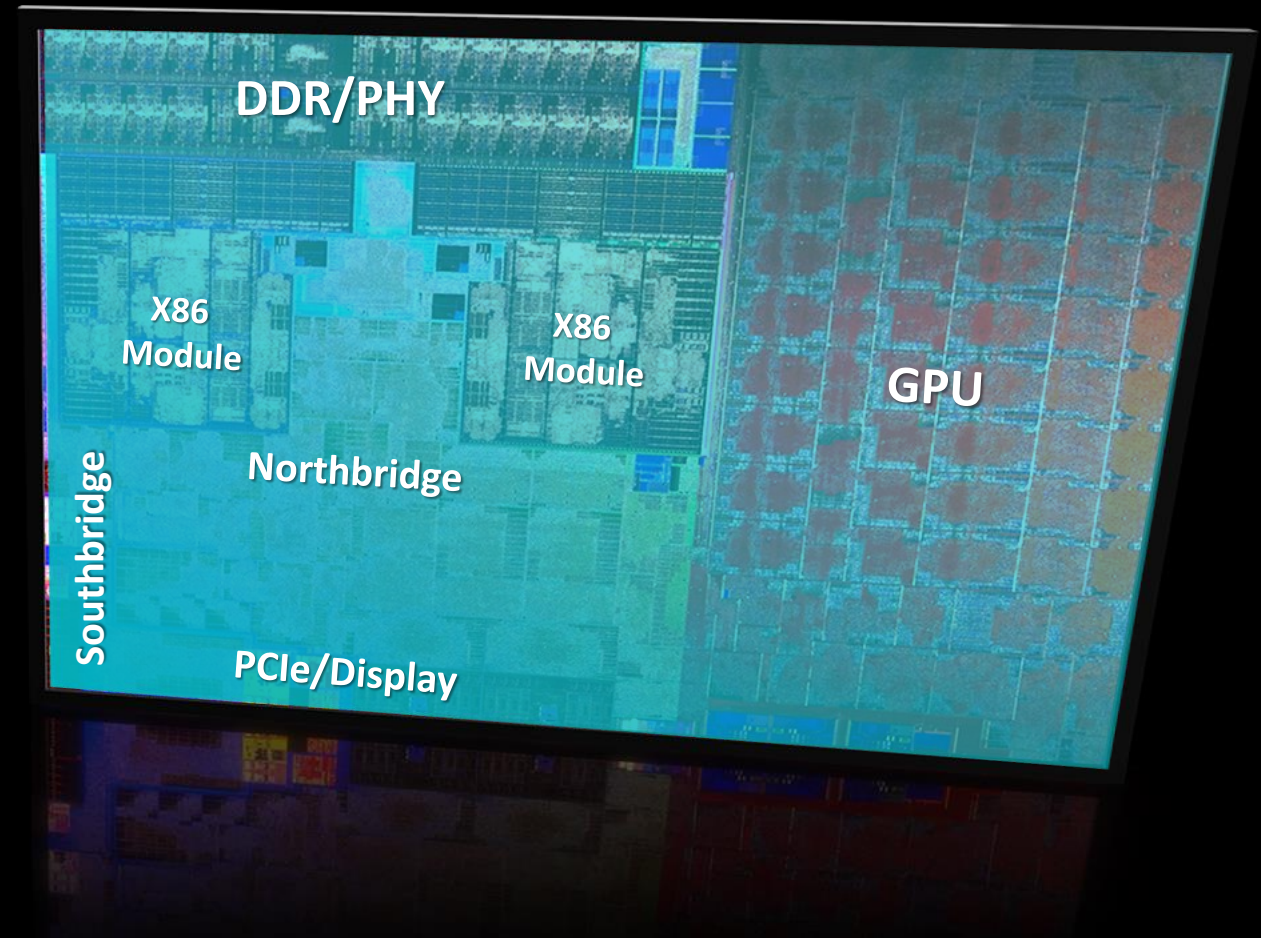
GUHAN KRISHNAN, DAN BOUVIER, LOUIS ZHANG, PRAVEEN DONGARA



## DESIGN GOALS

- ▲ Deliver superior performance, battery life and user experience for notebook and convertible form factors
- ▲ Deliver this energy efficiency gains in a mature, cost effective 28nm process node

## AMD 6<sup>TH</sup> GENERATION A-SERIES PROCESSOR



# 6<sup>TH</sup> GENERATION AMD A-SERIES PROCESSOR: “CARRIZO”

## SOC BLOCK DIAGRAM



### MAXIMUM COMPUTE PERFORMANCE

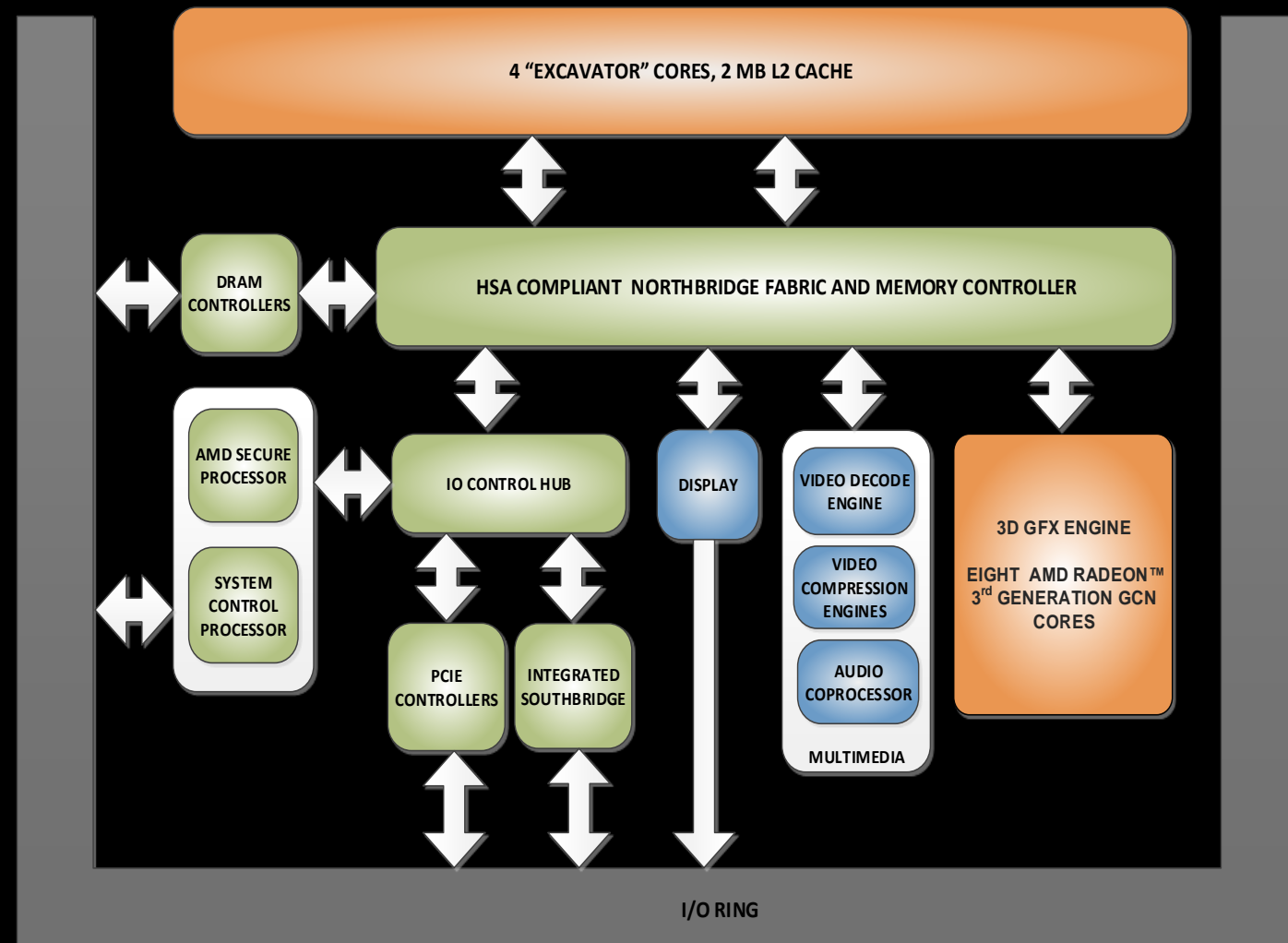
- 12 compute cores
  - 4 “Excavator” CPU cores
  - 8 GCN GPU cores
- HSA enabled

### ENHANCED USER EXPERIENCES

- HEVC/H.265 decode
- 3 display heads with UHD support
- Integrated security coprocessor

### HIGH PERFORMANCE CONNECTIVITY

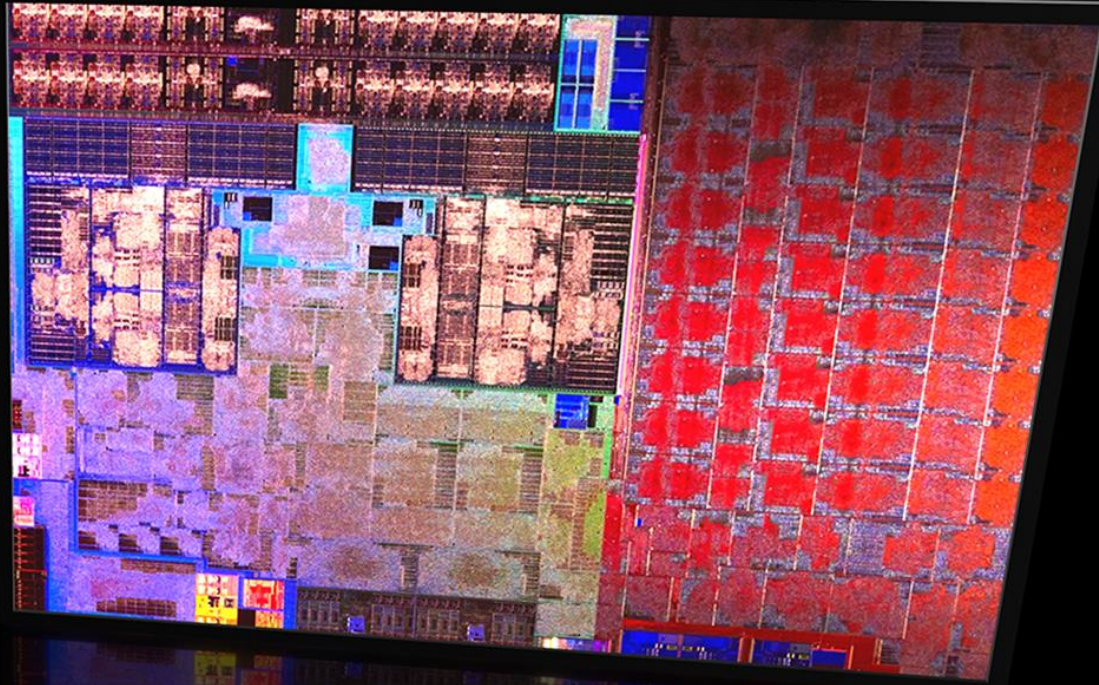
- 128 bits DDR3
- PCI-Express® Gen3 x8 for discrete graphics upgrade
- Integrated Southbridge



# 6<sup>TH</sup> GENERATION AMD A-SERIES PROCESSOR: “CARRIZO”



## ▲ OVERALL DIE STATISTICS



**Die size:** 250.04mm<sup>1</sup>

**Transistor count:** 3.1 billion

**Process:** 28nm Bulk high density

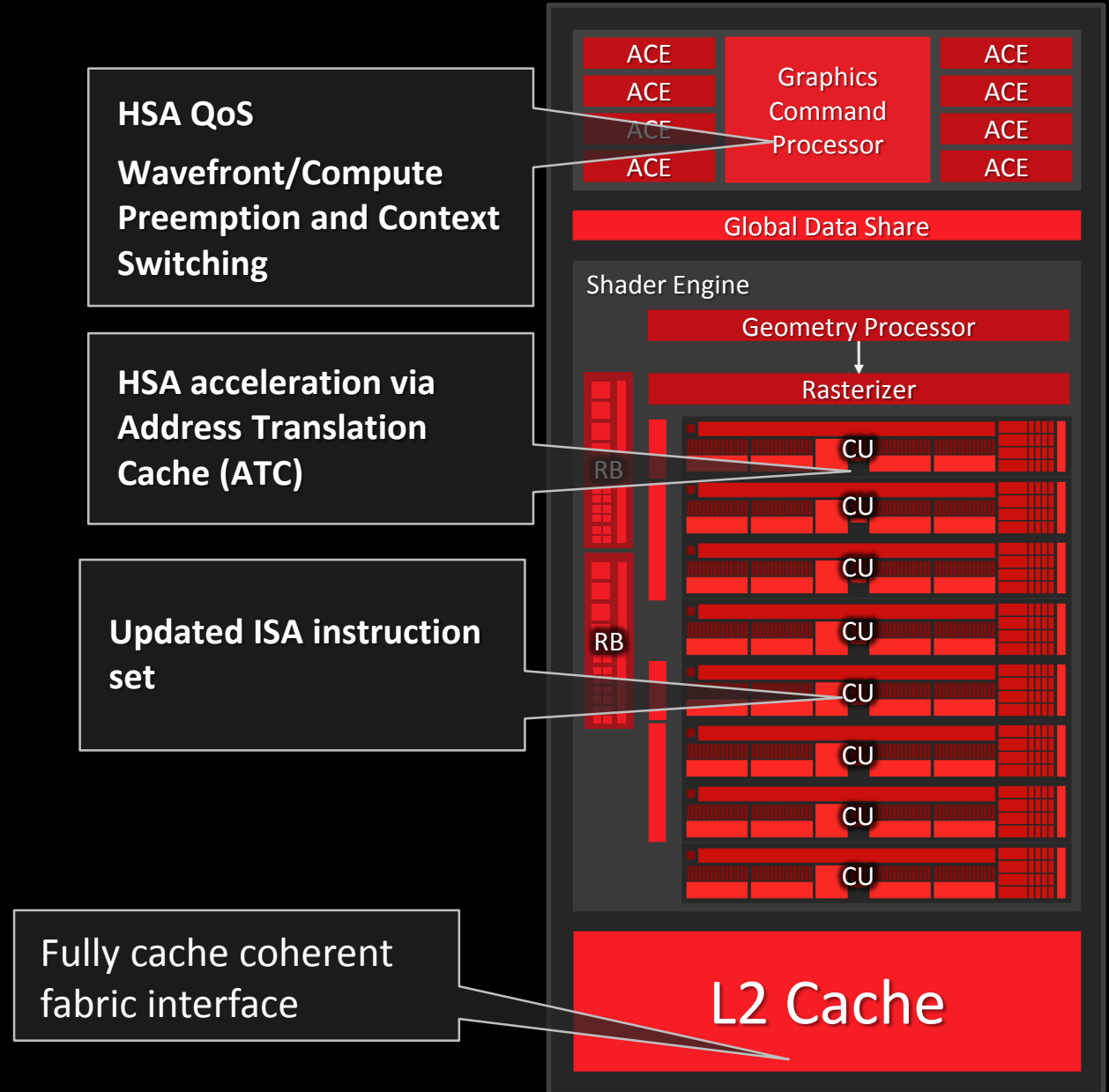
29% density increase over  
previous 28nm A-series APU<sup>1</sup>



# 6<sup>TH</sup> GENERATION A-SERIES PROCESSOR GRAPHICS ENGINE OVERVIEW



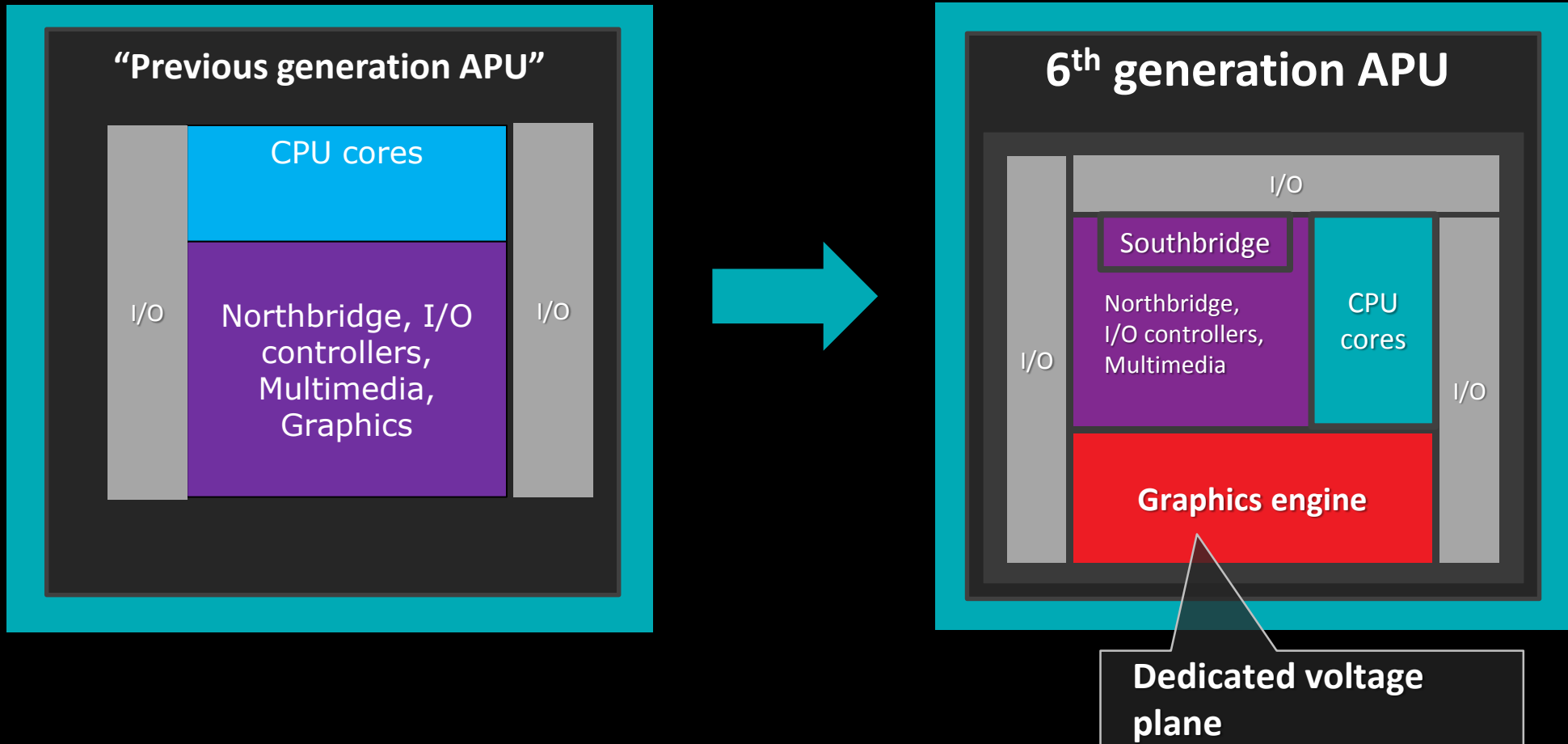
- ▲ Eight 3<sup>rd</sup> generation GCN cores
  - 819 GFLOPS
  - 512KB GFX L2 cache
  - DirectX 12 level 12
  - HSA acceleration and QoS
- ▲ Energy efficiency improvements
  - Graphic voltage island
  - Color compression
  - Low power implementation
  - Efficient power gating
  - GPU adaptive clocking



# GRAPHICS VOLTAGE ISLANDS



- ▲ Move Graphics core (33% of the overall die) to a separate voltage plane away from fabric and multimedia IP
- ▲ Independent voltage and frequency control based on Graphics application activity



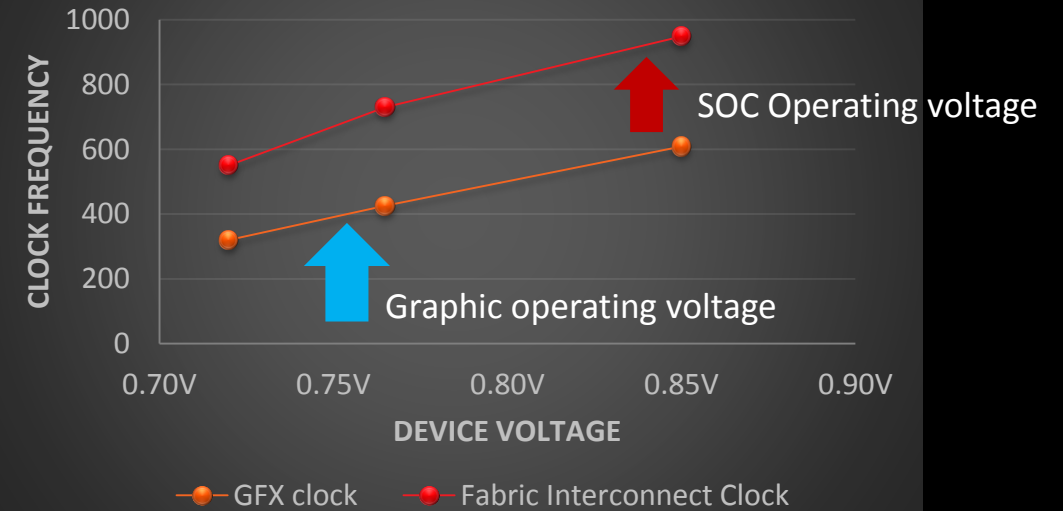
# GRAPHIC VOLTAGE ISLANDS

## MOTIVATION

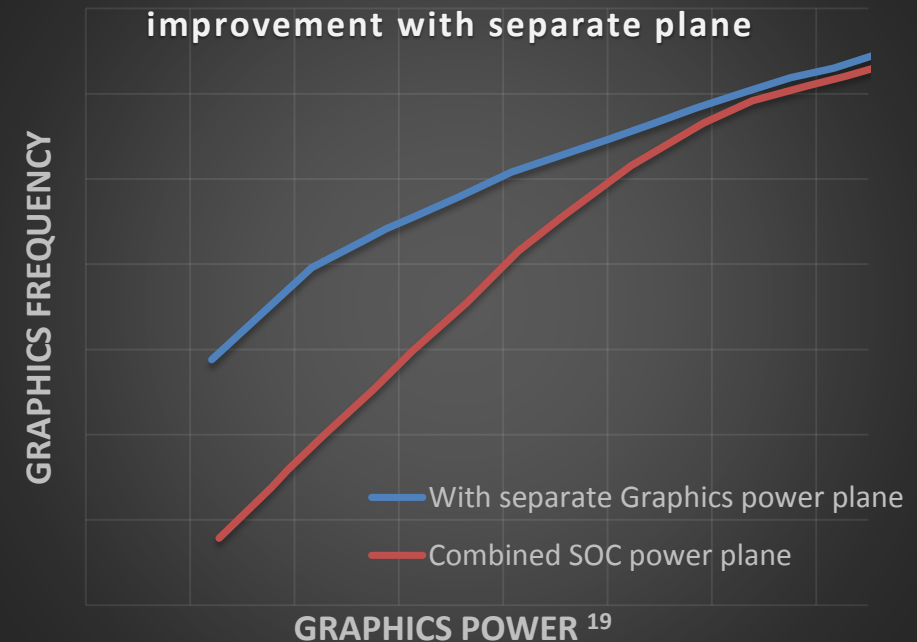
- ▲ Significant difference in steady state voltage levels between SoC and Graphic planes during PC gaming use case
  - Ability for the Graphics engine to operate at the voltage dictated by the application activity levels
- ▲ Intensive multimedia use cases require higher SoC voltage for fixed function engine operation
  - Thermally sustainable graphics operation is not possible for an unified plane in a 15W thermal dissipation envelope
- ▲ Augments other power gating within the graphics core
  - Idle desktop, productivity use cases



Steady State voltage difference during PC gaming



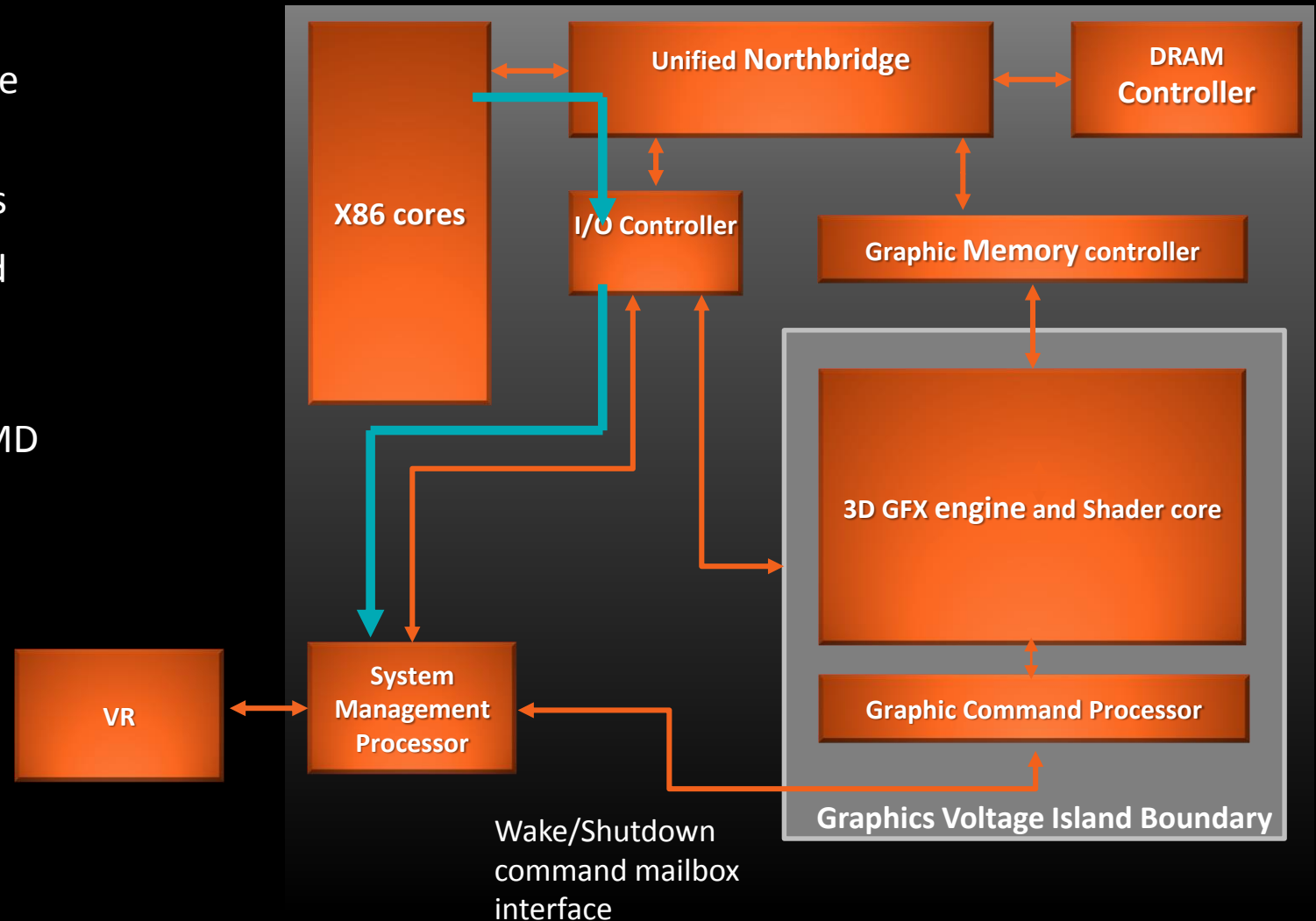
Power constrained frequency improvement with separate plane



# GRAPHIC VOLTAGE ISLANDS

## IMPLEMENTATION

- ▲ Wake path activated by “doorbell” write to Graphics core (shown in blue)
- ▲ Shutdown based on “idle” time periods
- ▲ Wake/Shutdown sequencing controlled by system management processor
- ▲ Graphic voltage supplied by external voltage regulator (VR) controlled by AMD SVI 2.0 interface
- ▲ Voltage crossings hardware between domains
  - Less than 0.1% die area cost

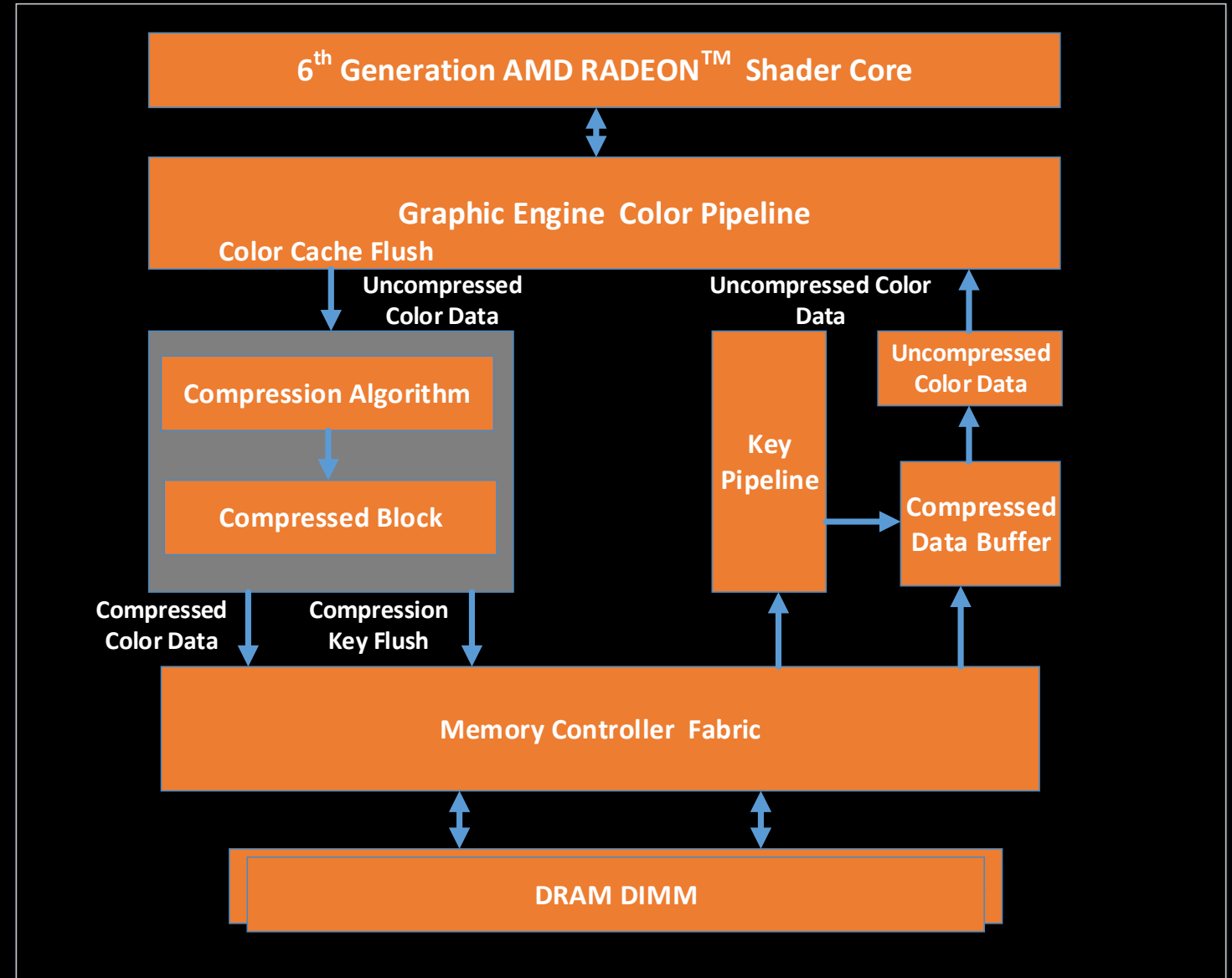




# GRAPHIC COLOR COMPRESSION

## MOTIVATION/IMPLEMENTATION

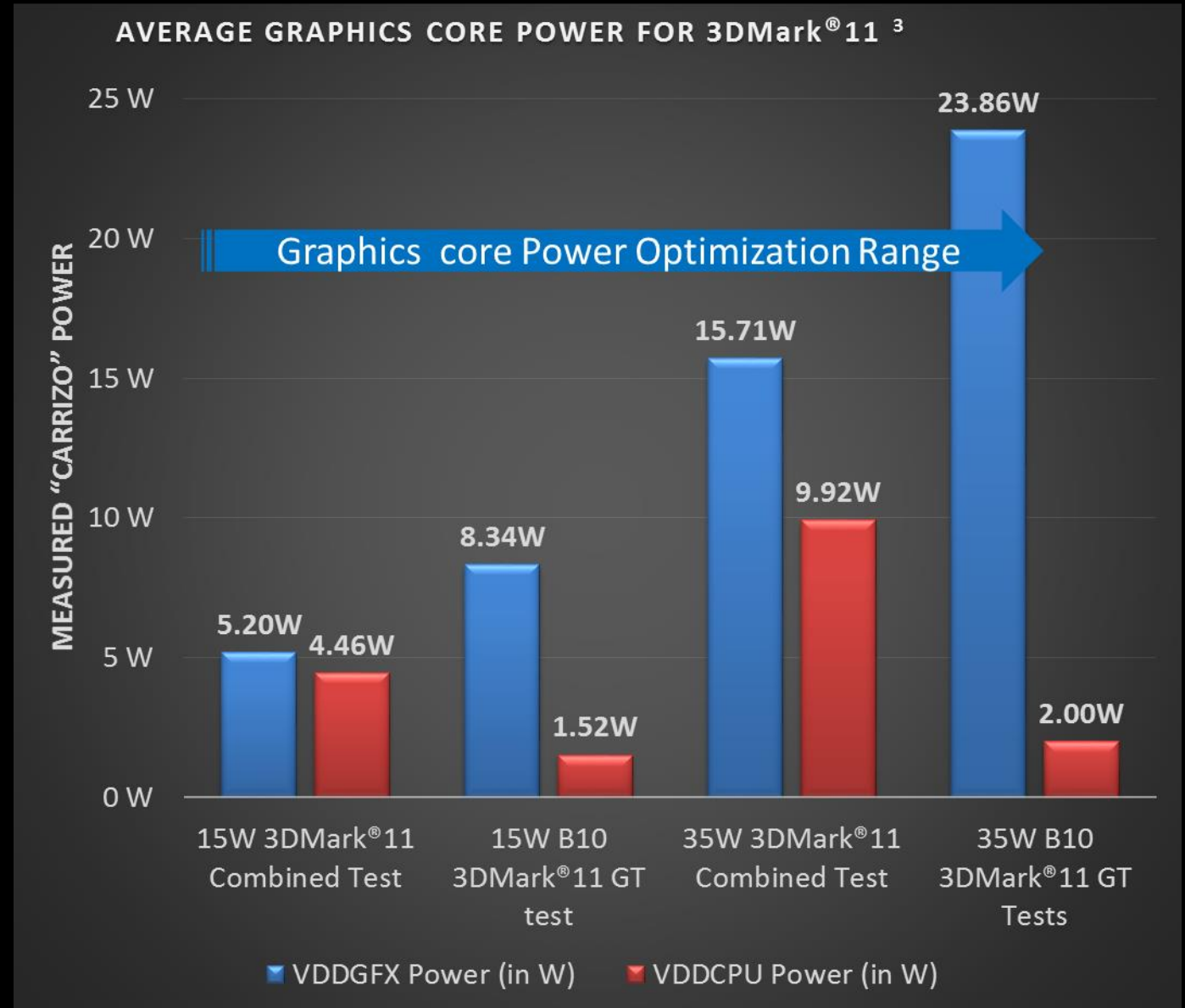
- ▲ Reduced DRAM bandwidth for Graphics render reduces the total power of the system
  - 40%<sup>19</sup> of DRAM accesses during typical PC gaming is color traffic
- ▲ Many PC systems are sold with only a single channel of memory populated
  - Higher pressure on DRAM bandwidth for Graphics render
- ▲ Graphics reads/writes compressed data
  - Compression during cache flush
  - Decompression on the read return
- ▲ Transparent to Graphics driver software
  - 5-7%<sup>2</sup> improvement on games for modest silicon area (0.2%)



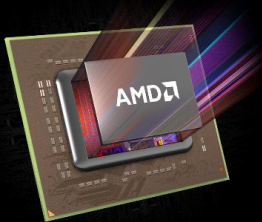
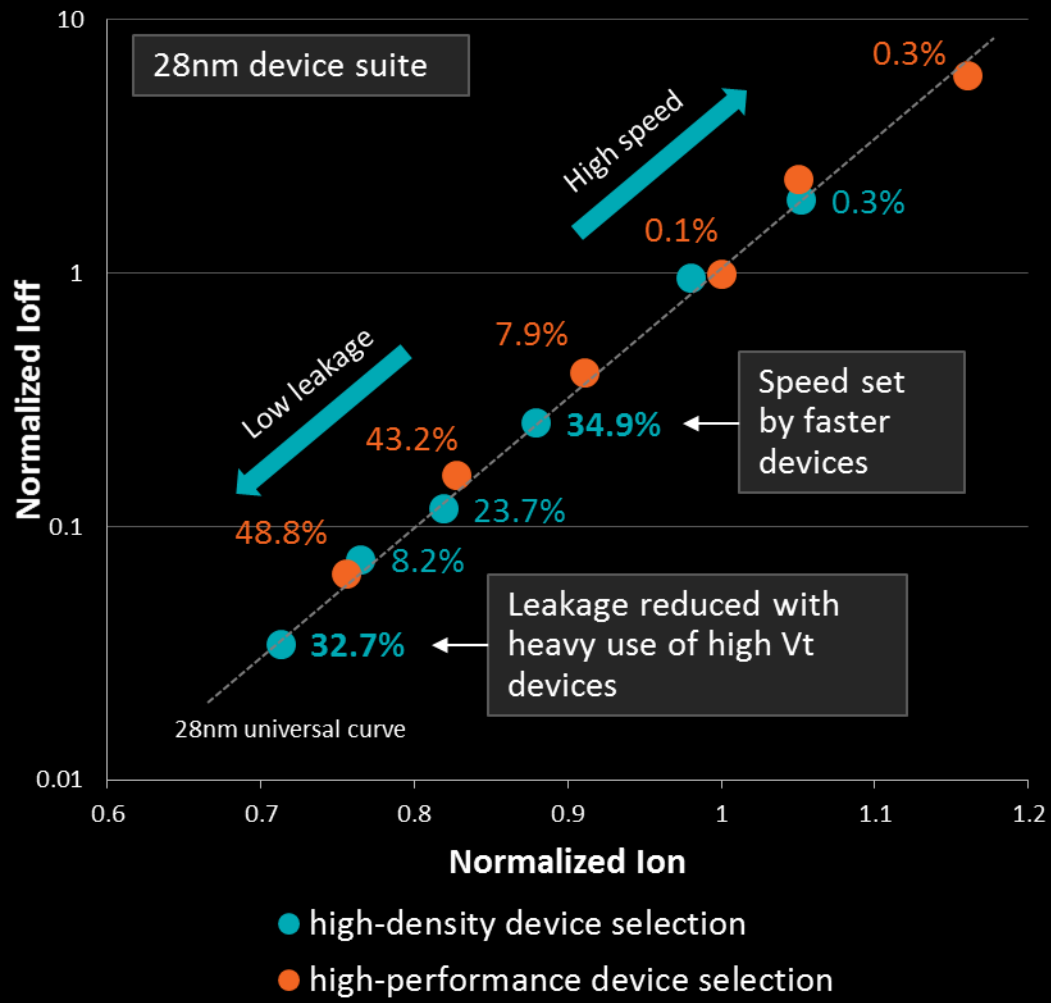
# LOW POWER OPTIMIZED GRAPHICS

## MOTIVATION – TARGETING OF THE GRAPHICS CORE

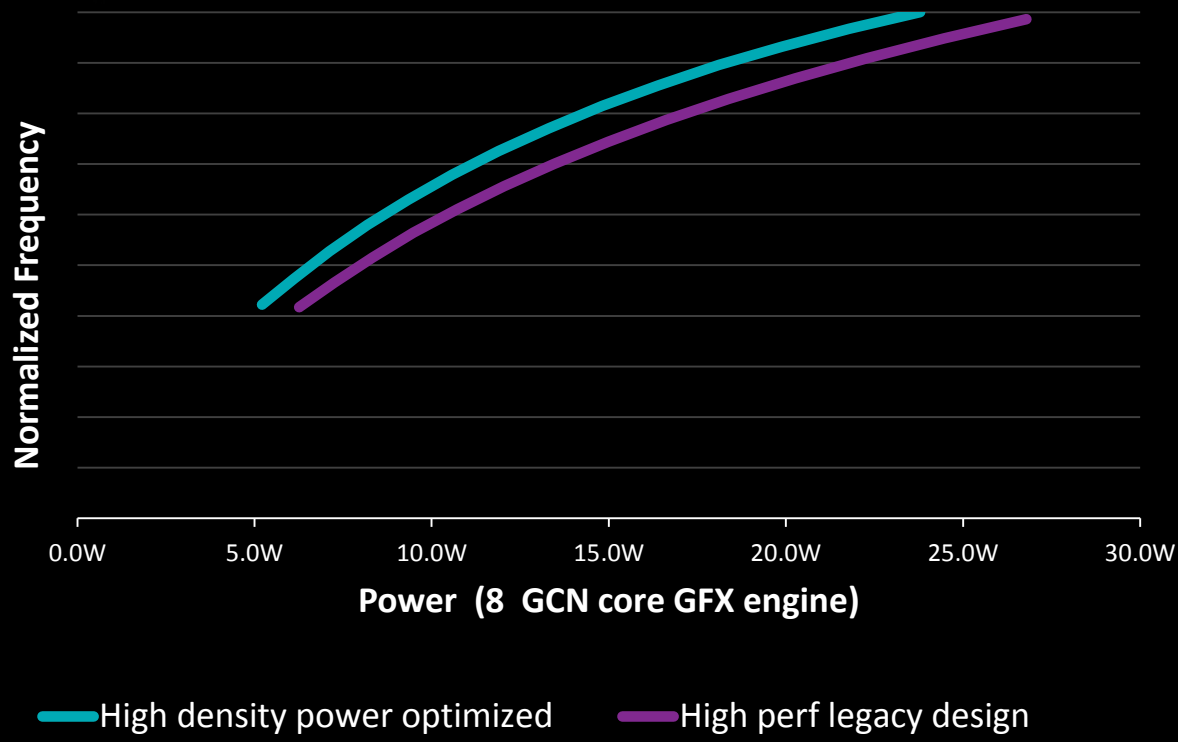
- ▲ Optimize graphics core operation for a 15W SOC
- ▲ Target 28nm devices to lower leakage device by 2.5x for a 10% loss of drive strength
  - Reduce leakage at Vmin to give up some performance at Vmax
- ▲ Enables maximum parallel operation of graphics core for a power range 5W-25W
  - Allows all 8 GCN compute core to operate in Vmin for a 15W SoC, 33% more than a Kaveri 15W SoC



# LOW POWER OPTIMIZED GRAPHICS



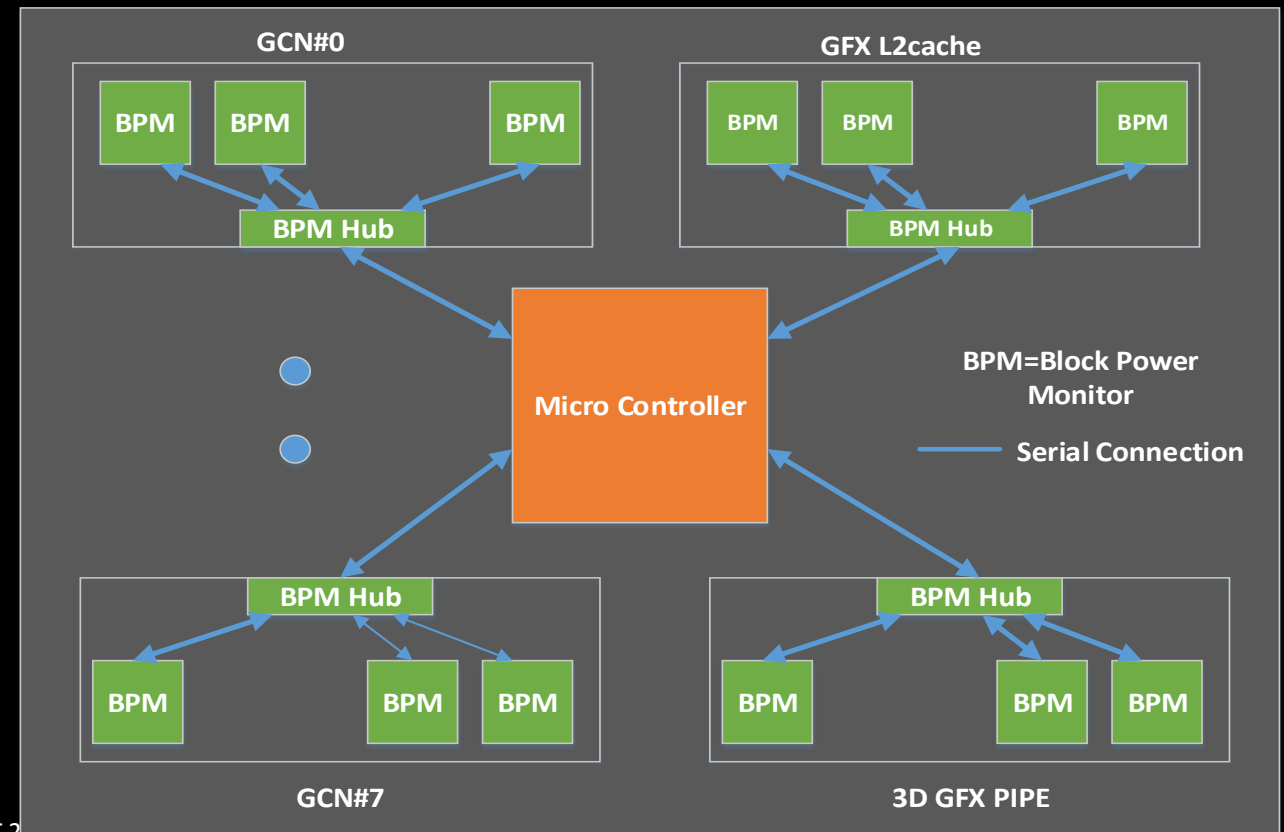
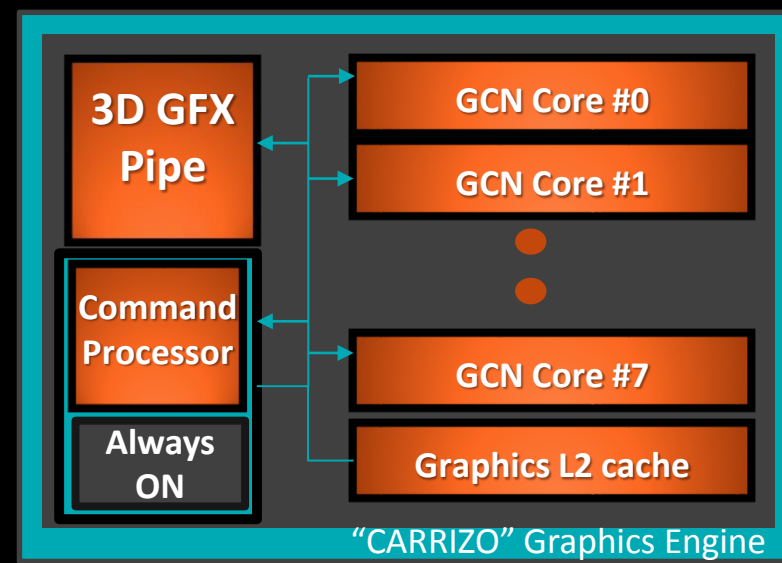
*18% leakage reduction and timing with faster RVT devices enables 10% higher frequency at same power level, or up to 20% lower power at same frequency<sup>1</sup>*





# GRAPHIC ENGINE POWER GATING

- ▲ Several sub-power gating domains within the graphics engine to augment coarse/medium/fine grain clock gating
  - Per GCN core power gating
  - 3D graphics pipe power gating
- ▲ Power gating is controlled by micro controller running on “always ON” domain in the command processor
  - Aggressive wide enablement of number of cores as long as we are within infrastructure limits
  - Up/down control by firmware in conjunction with activity and thermal

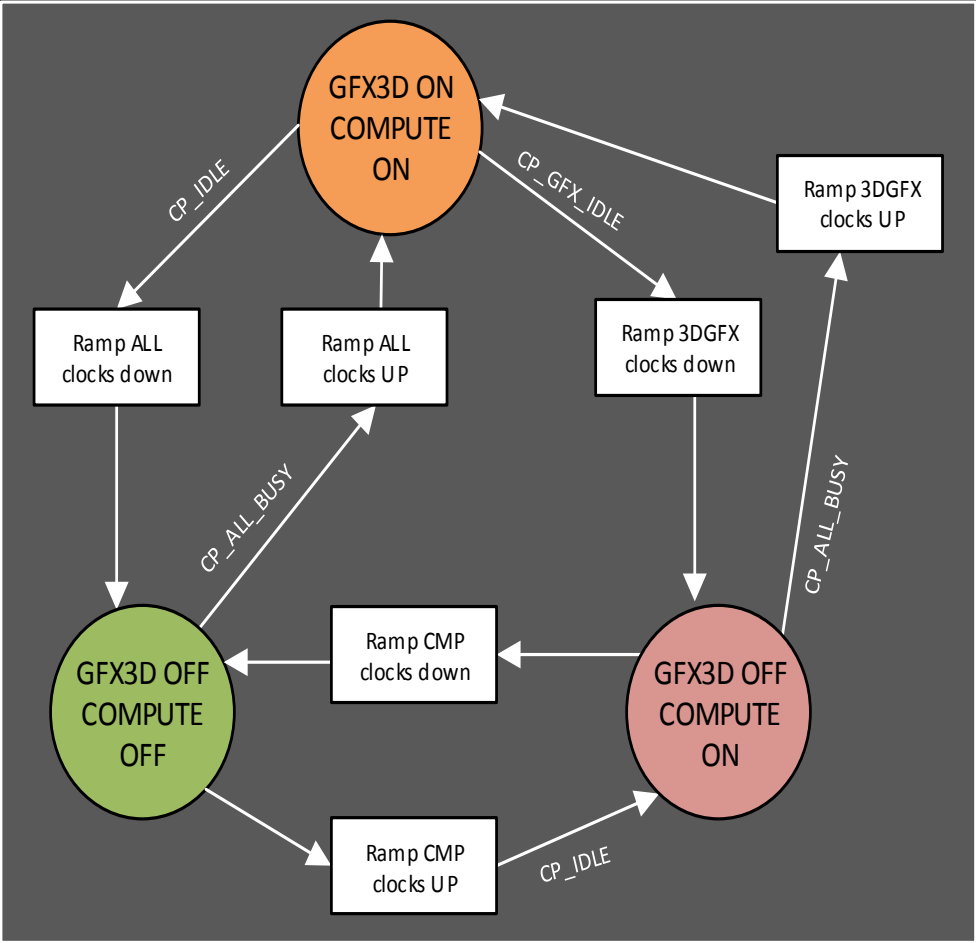


# GRAPHICS POWER GATING

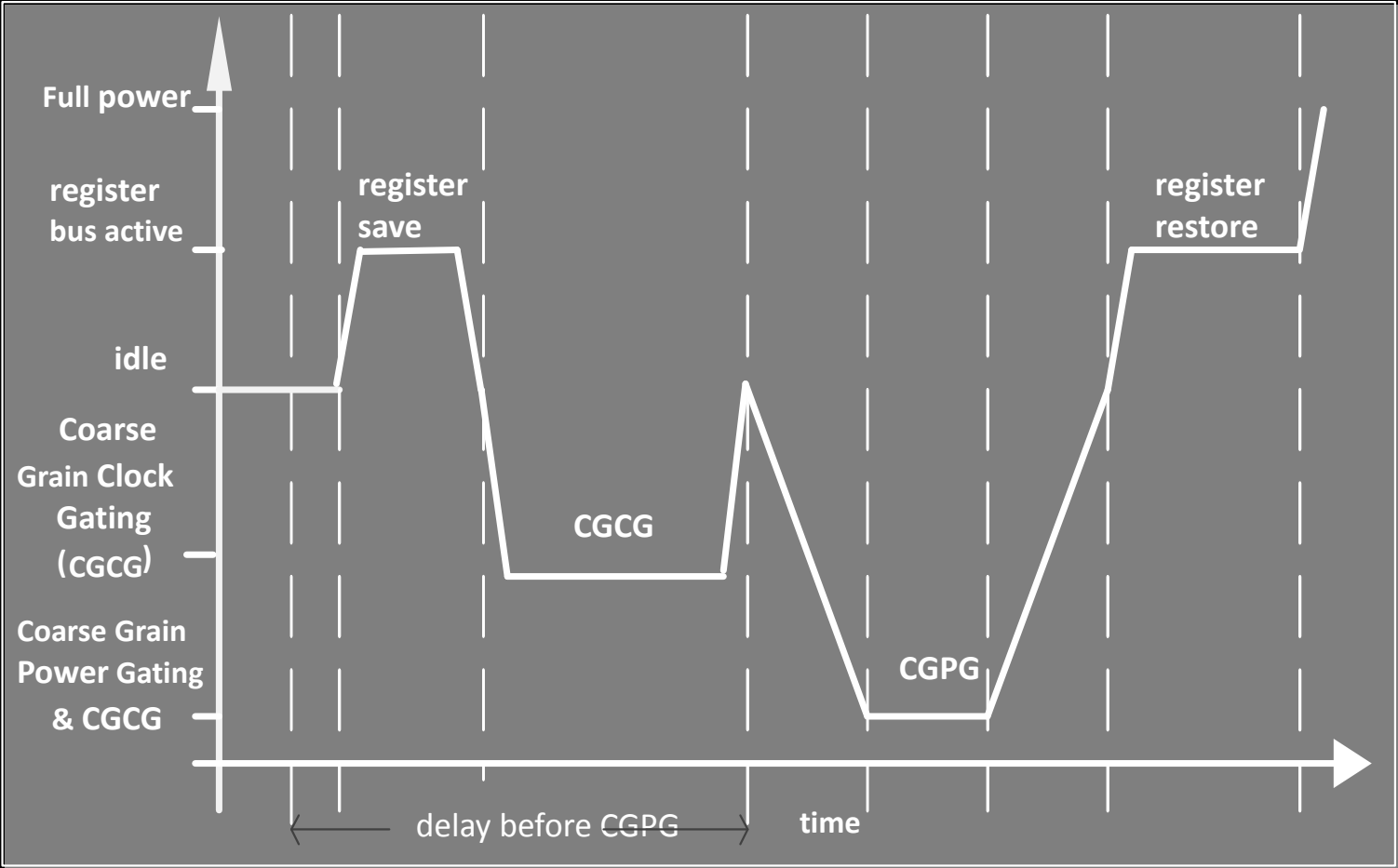
CONTROLLED STEP DOWN IN POWER OVER TIME



POWER GATING STATE MACHINE



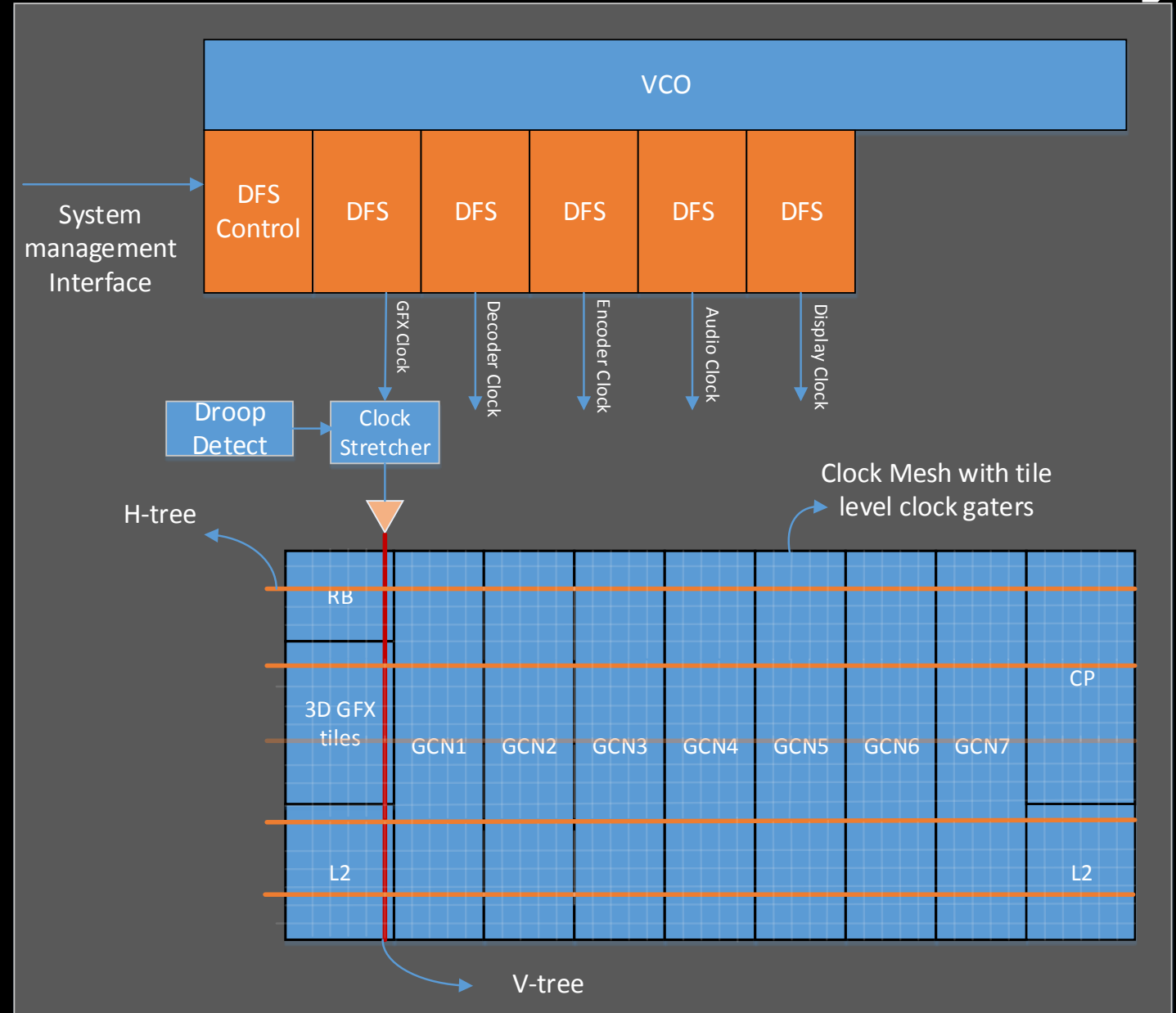
POWER GATING FLOW



# GRAPHIC ENGINE CLOCKING

## EFFICIENT CLOCK DISTRIBUTION AND GATING

- ▲ Digital frequency synthesizer (DFS) that generates multiple discrete frequencies from a single VCO
  - Root clock gating
  - Disabling VCO and bypassing with low speed fixed clocks
- ▲ Custom clock distribution with 5 H-tree, 1 V-tree feeding a clock mesh
  - Coarse and medium grain gating
  - Dynamic clock gating based on load balancing
- ▲ Adaptive clocking mitigates frequency loss due to voltage droop events caused by  $di/dt$ 
  - Dynamically detect droop and stretch clocks
  - Eliminates higher voltage to address droop associated with the worst case pattern



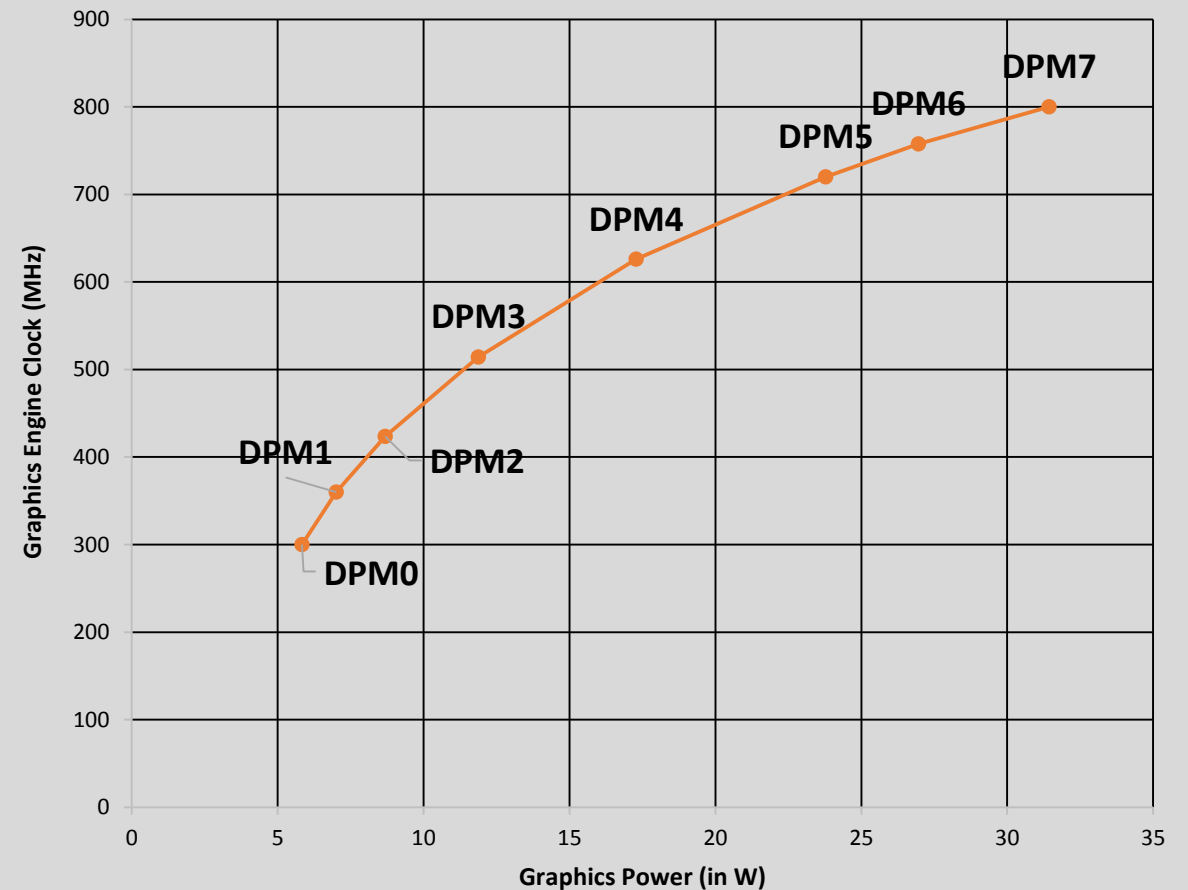


# GPU POWER/PERFORMANCE MANAGEMENT



- ▲ 8 discrete performance states representing 8 discrete voltage, frequency, and power states that graphics can operate
  - Chosen to cover  $V_{\min}$  to  $V_{\max}$  operation
  - Optimized for high, medium, and low CaC operation
- ▲ Algorithm that quickly responds to graphics activity by increasing SCLK frequency
  - Graphics activity >0% will cause SCLK to go to the highest frequency
  - Graphics inactivity (0% busy) will cause SCLK to go to the lowest frequency
  - Waterfall thresholds of CaC cause up/down
  - Applications will settle on a steady state power budget based on overall chip activity
  - Additional response filtering in battery mode

Graphics application power versus engine clock



## GRAPHICS PERFORMANCE

- ▲ Optimized for the 15W SOC TDP
  - Designed to enable increased frequency for up to 18% additional performance
  - Designed to enable utilization of the full 8 graphics cores rather than just 6 on previous generation APU, for an additional 20+%
  - Improvements in execution efficiency contribute another ~10%
- ▲ Total increase of up to 65% in the key graphics 3DMark<sup>®</sup> 11 benchmark over previous generation<sup>4</sup>

6<sup>th</sup> Generation APU "Carrizo" 3DMark<sup>®</sup> 11-P Performance vs. previous generation APU code named "Kaveri"<sup>4</sup>

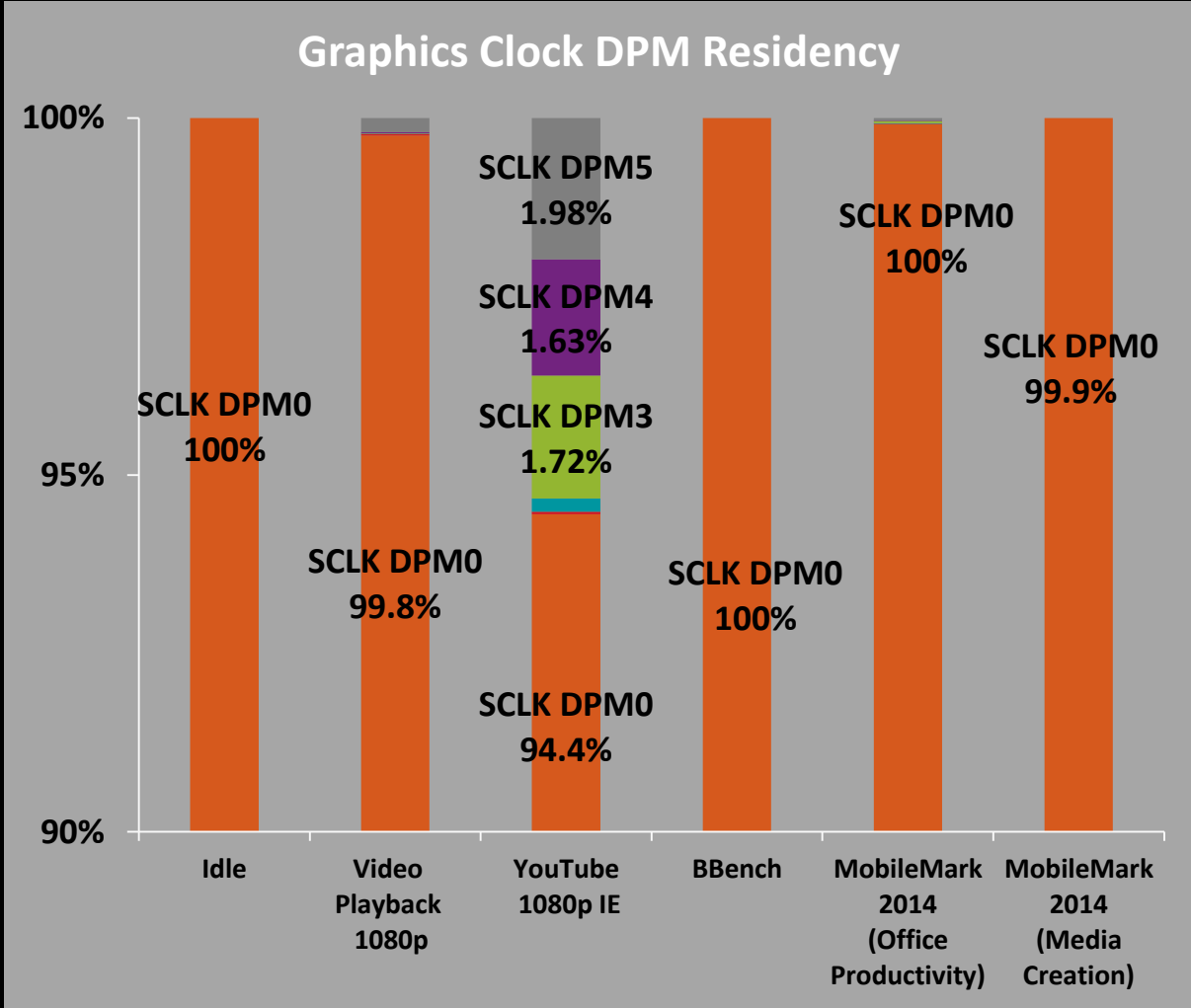
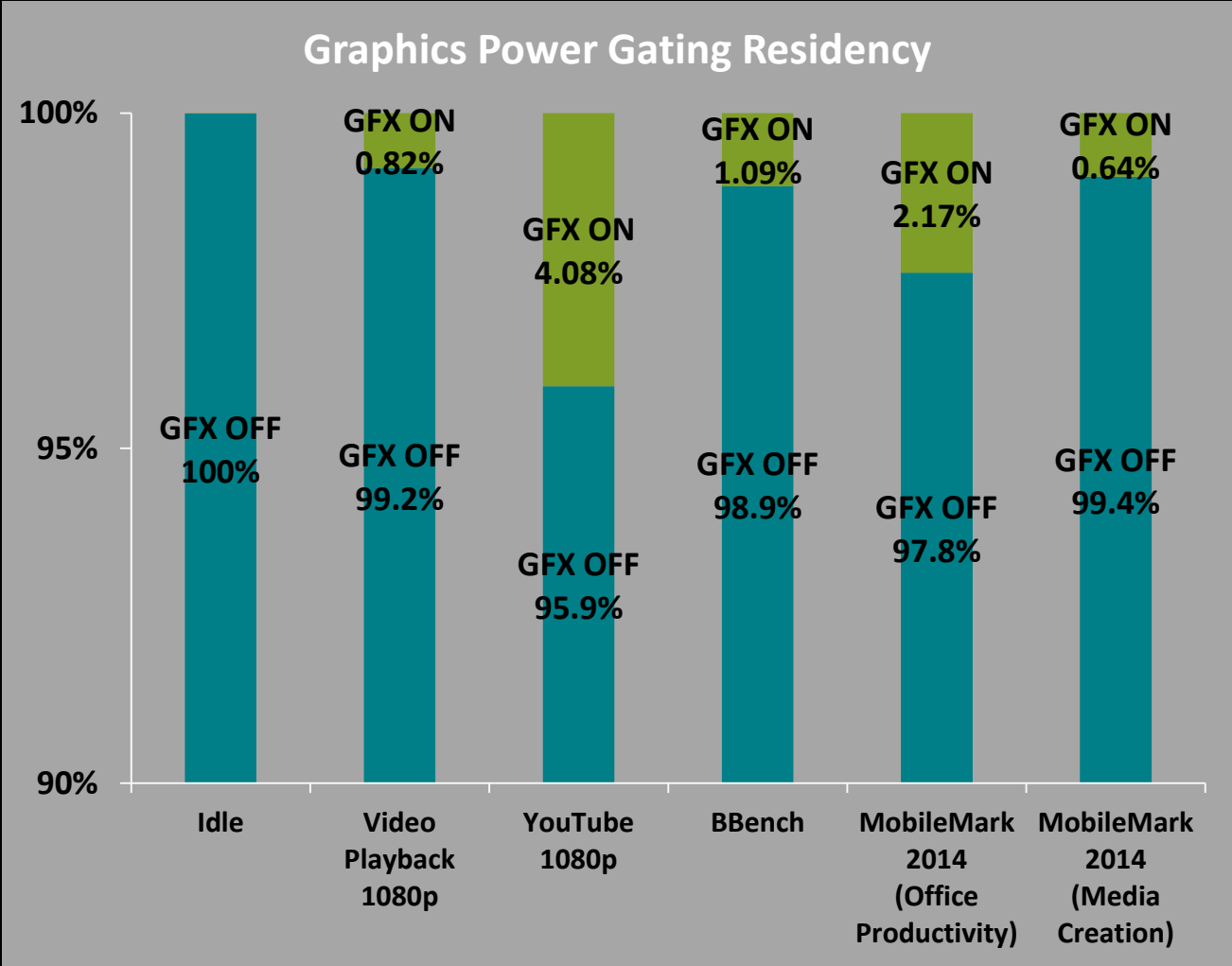


# PUTTING IT ALL TOGETHER



## BATTERY USE CASES

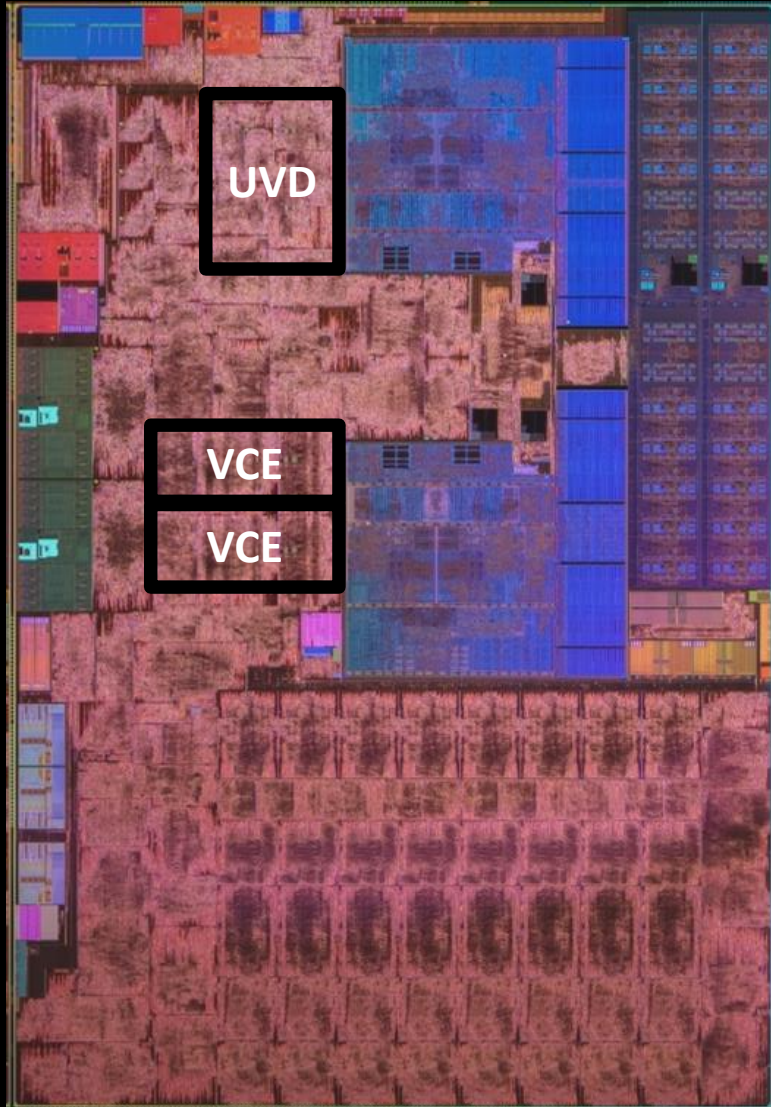
▲ High 90's power gating and low clock state residencies on common mobile use cases





# VIDEO IP OVERVIEW

SIGNIFICANT AREA INVESTMENT – 2.8X MORE THAN “KAVERI”



## VIDEO DECODER

- ▲ HEVC/H.265 decode
- ▲ Native 4K H.264 decode
- ▲ Support for hardware overlay hardware

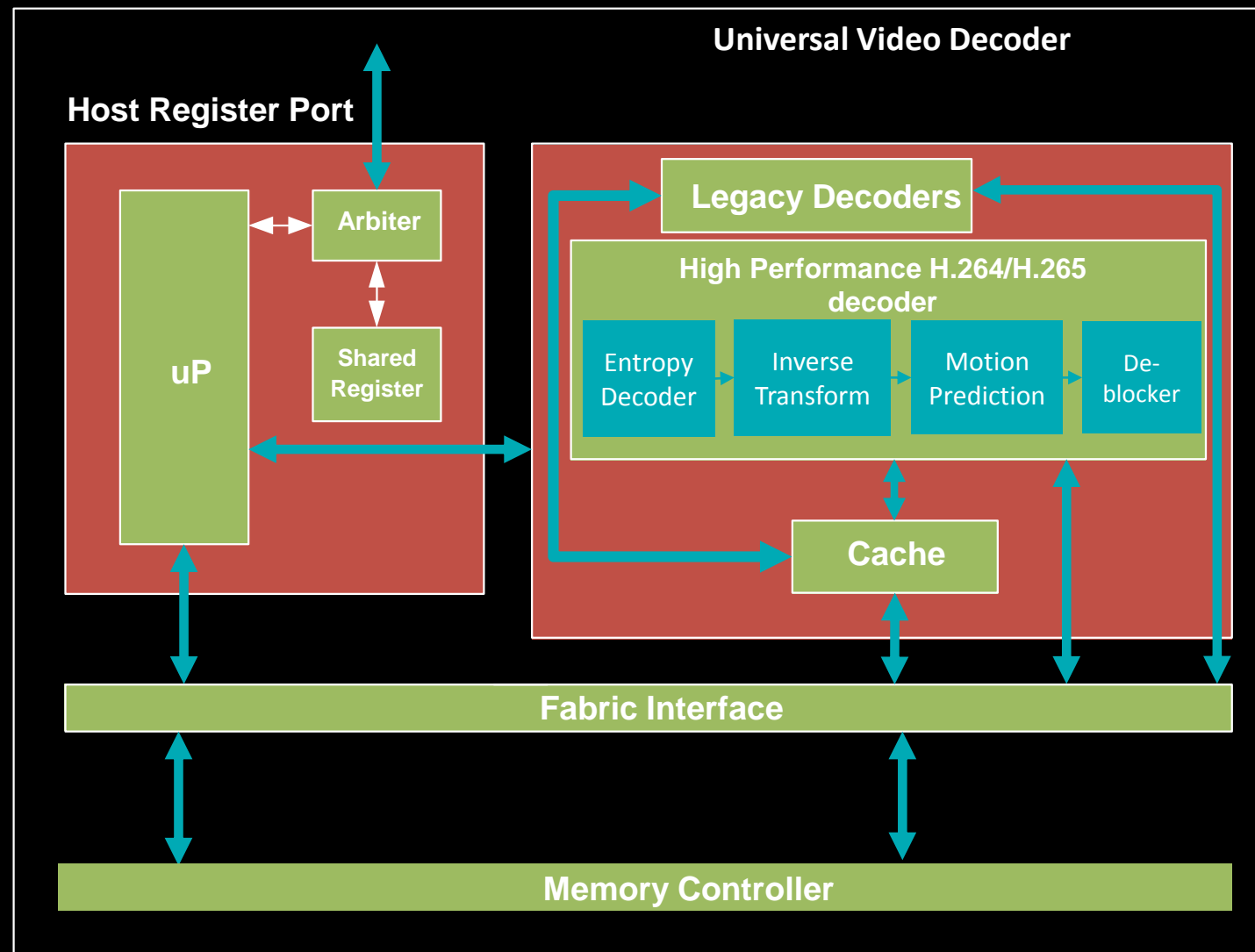
## VIDEO COMPRESSION ENGINE

- ▲ Dual compression engines
- ▲ Up to 3.5X improvement on 1080p transcode<sup>1</sup>
- ▲ Full 1080p@60 Wireless MiraCast

# VIDEO DECODER ARCHITECTURE OVERVIEW



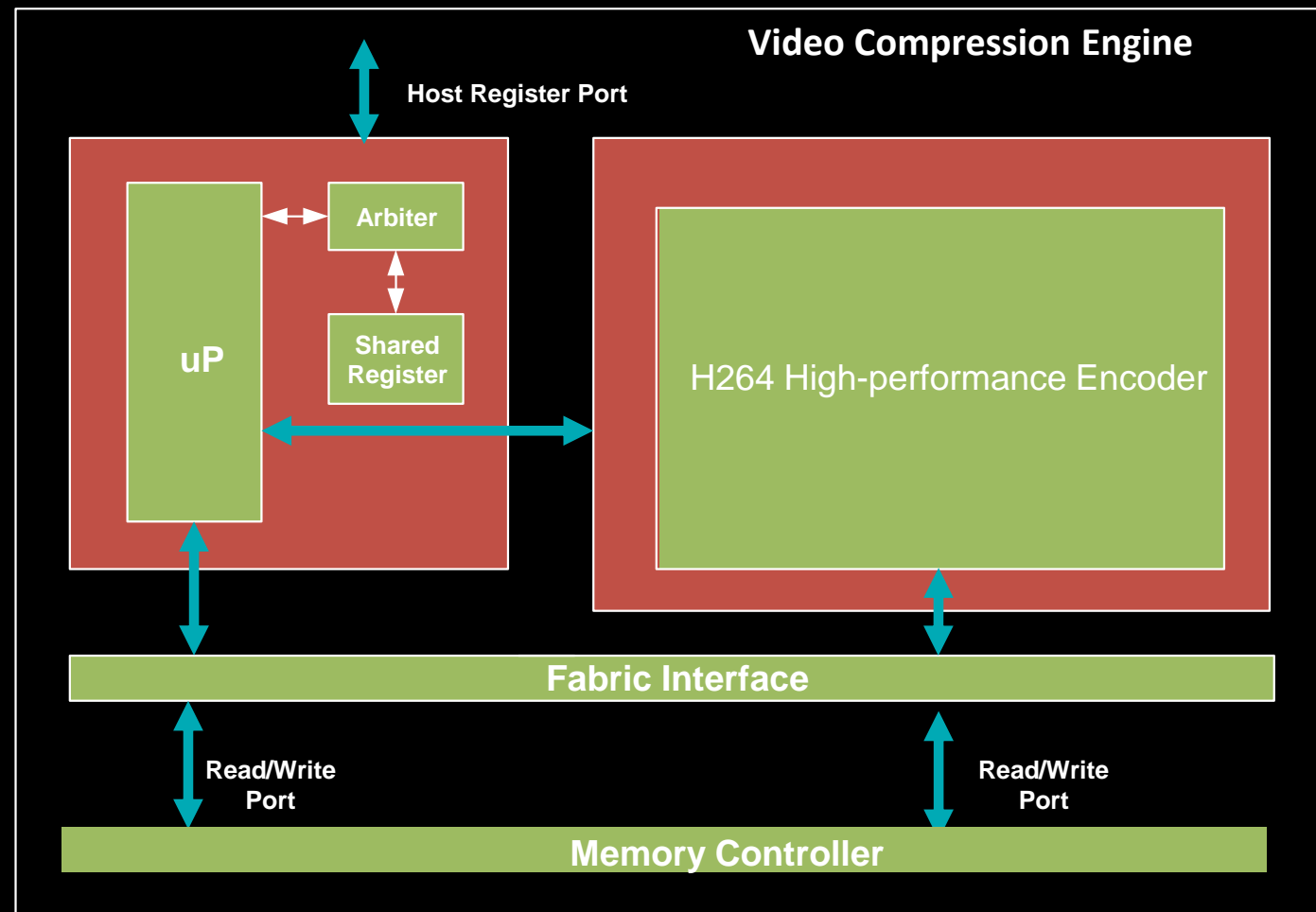
- ▲ 6<sup>th</sup> generation A-series processor houses a redesigned video decoder to deliver fast decode times
  - 3X bigger than KV
  - 350%<sup>6</sup> reduction in decode time for 1080p video content<sup>6</sup> when compared to KV
- ▲ Enhances user experience by supporting hardware decode HEVC/H.265 format
  - 2X lower power than S/W decode<sup>5</sup>
- ▲ Improved memory efficiency with more internal storage/caching
- ▲ Faster decode times allows faster lowest power state for the SoC and DRAM



# VIDEO ENCODER ARCHITECTURE OVERVIEW



- ▲ 6<sup>th</sup> generation A-series processor houses a dual H.264 encoder instances
  - Each VCE instance having a twin front end pipeline to handle encoding jobs
  - 4K H.264 encode
- ▲ Return on investment for encoding area
  - Up to 169 fps 720p to 720p<sup>7</sup>
  - Up to 78 fps 1080p to 720p/1080p<sup>7</sup>
  - Represents 350% improvement over KV
- ▲ Encode throughput capable of supporting 1080p@60, 4K@30 Wireless Display

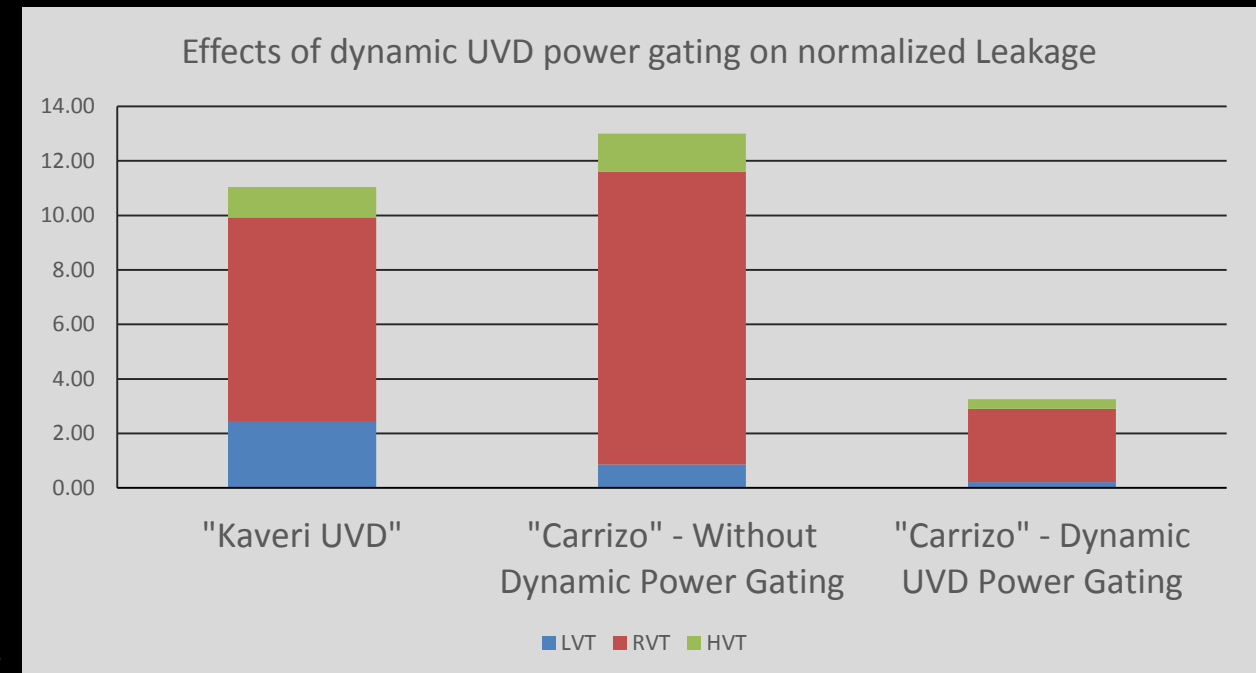
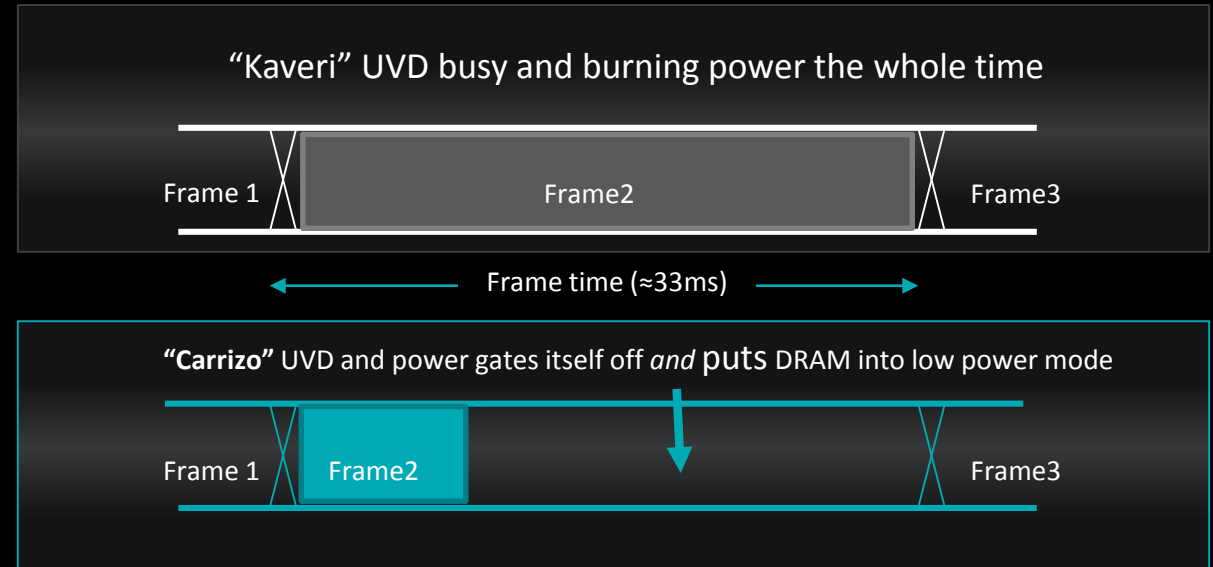




# DYNAMIC UVD POWER GATING



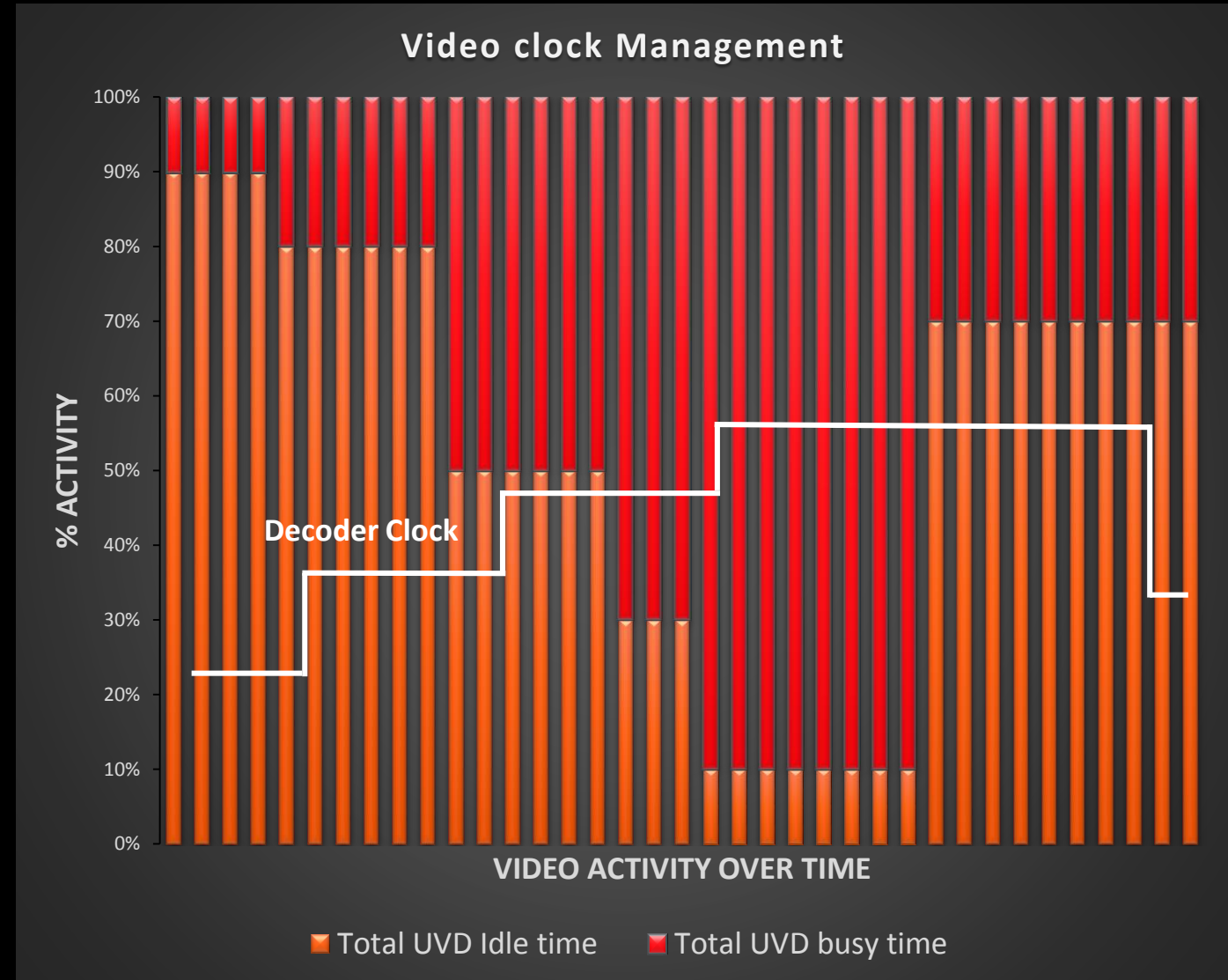
- ▲ Dynamic inter frame power gating controlled by microcontroller firmware
  - Pipeline idle detection enables header/footer power gating of the entire IP
- ▲ Dynamic power gating along with low power hardening of the video decoder enables CZ to negate the bigger video decoder needed for H.265 offload
  - ~3X better than KV in net leakage profile<sup>21</sup>



# DYNAMIC VIDEO DECODER CLOCK MANAGEMENT

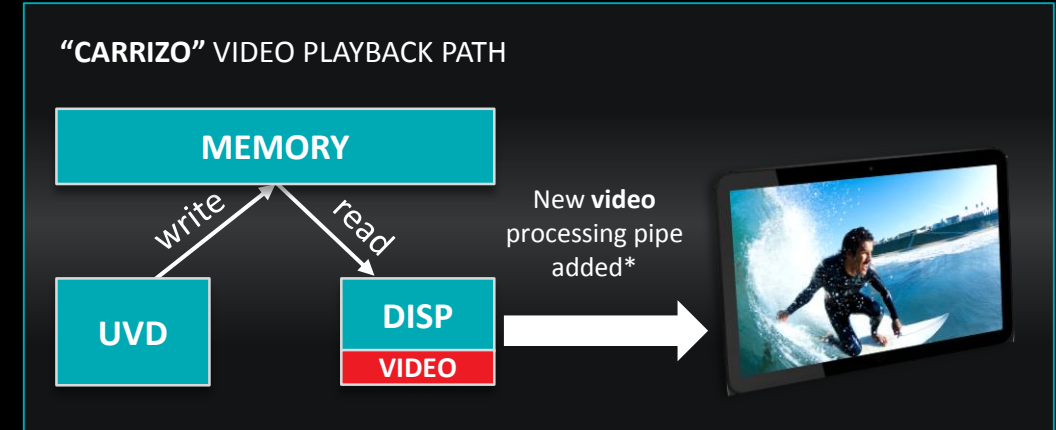
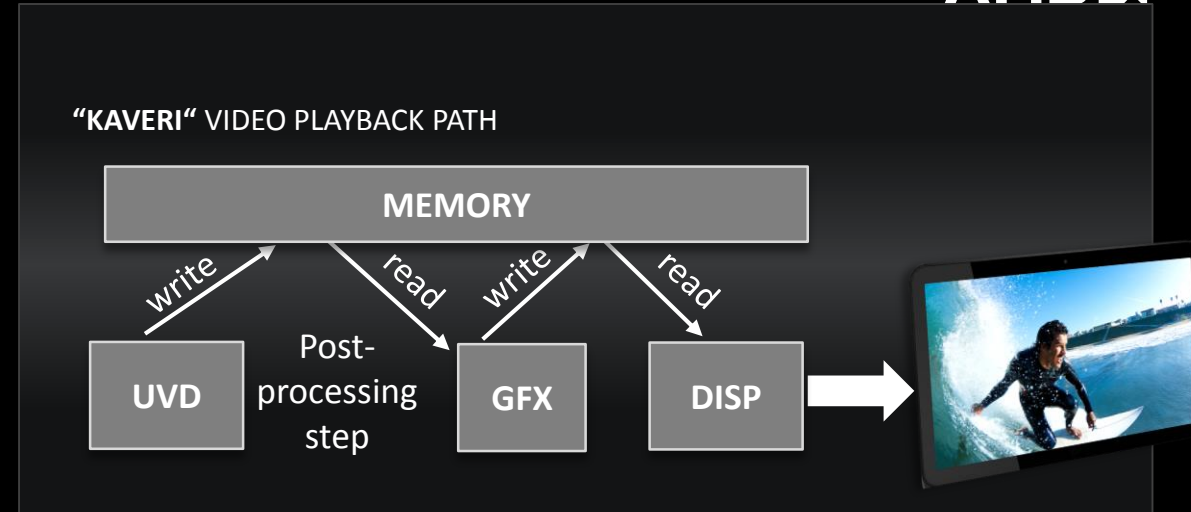


- ▲ Adjust video decoder clock dynamically for power savings based on frame decode time
- ▲ Calculate running average of decode time of last <X> frames and idle times of the last <Y> frames
  - Use the values from last X number of “busy” times for bumping up the clock
  - Use the values from the last Y number of “idle” times frames to slow down the clock
- ▲ Firmware predicts performance increase/decrease before requesting a clock change



# OPTIMIZED VIDEO PLANE

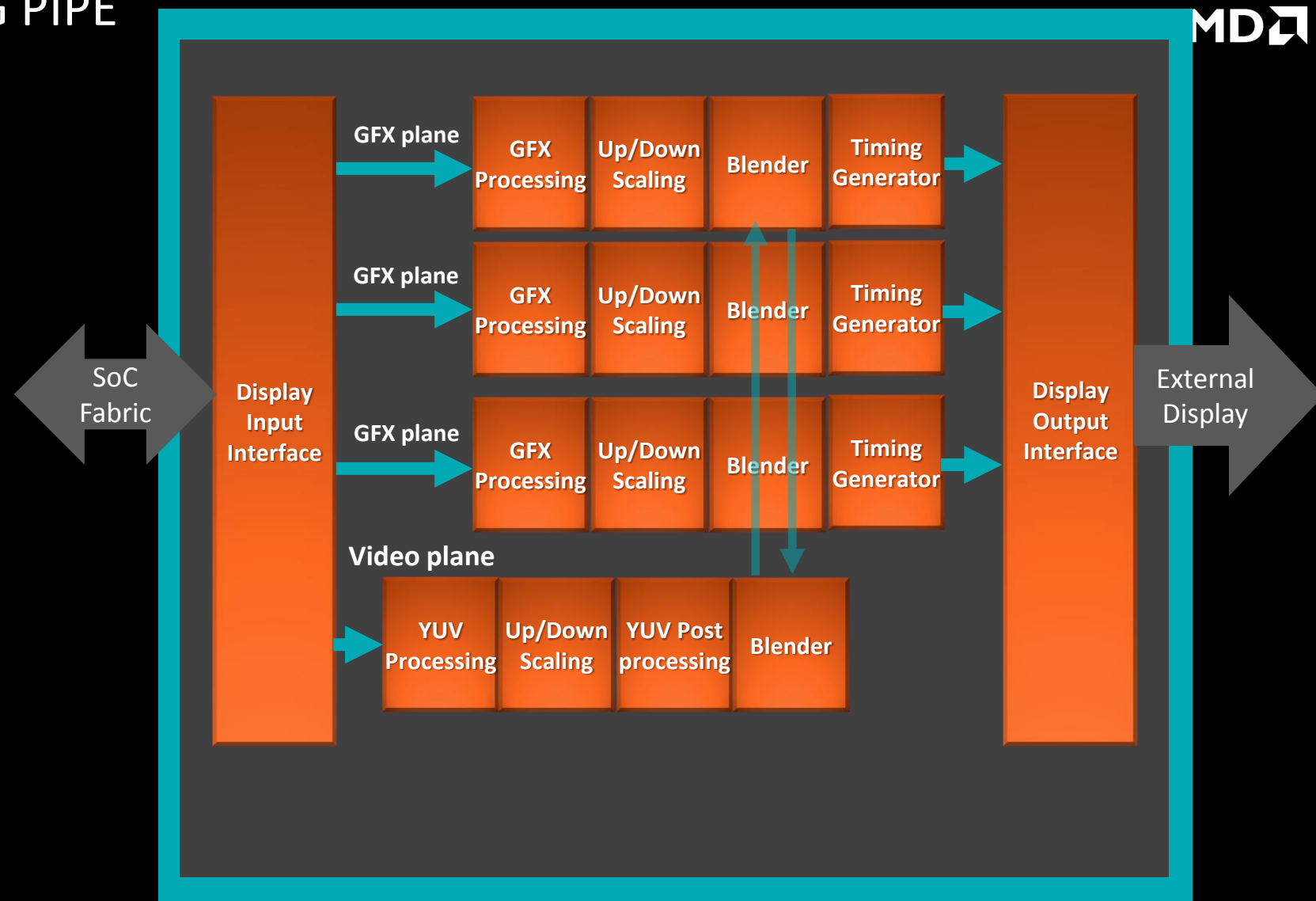
- ▲ Traditionally the GPU shader engine is required to scale and process images during video playback
  - This consumes a lot of power, utilizes the “big” GPU poorly
  - Additional bandwidth to DRAM hops burns SoC and platform power
- ▲ Operating systems allow a dedicated video plane for hardware to process video efficiently
  - Popularly known as “multi-plane overlay” or MPO
- ▲ Small amount of dedicated logic added to display control engine to process video
  - Net leakage added dwarfs the benefits of reducing GFX post-processing and DRAM power



\*Path cannot be enabled in all cases, there will be cases where software will enable the Graphics CU's for post-processing

# DISPLAY VIDEO PROCESSING PIPE

- ▲ Hardware in display engine to process YUV data
- ▲ Reuses the existing SoC fabric interfaces to fetch video data from memory
- ▲ Blender functionality allows merge with any of the three primary display pipes
- ▲ Hardware has restrictions on the scaling and rotation of the video it can support natively

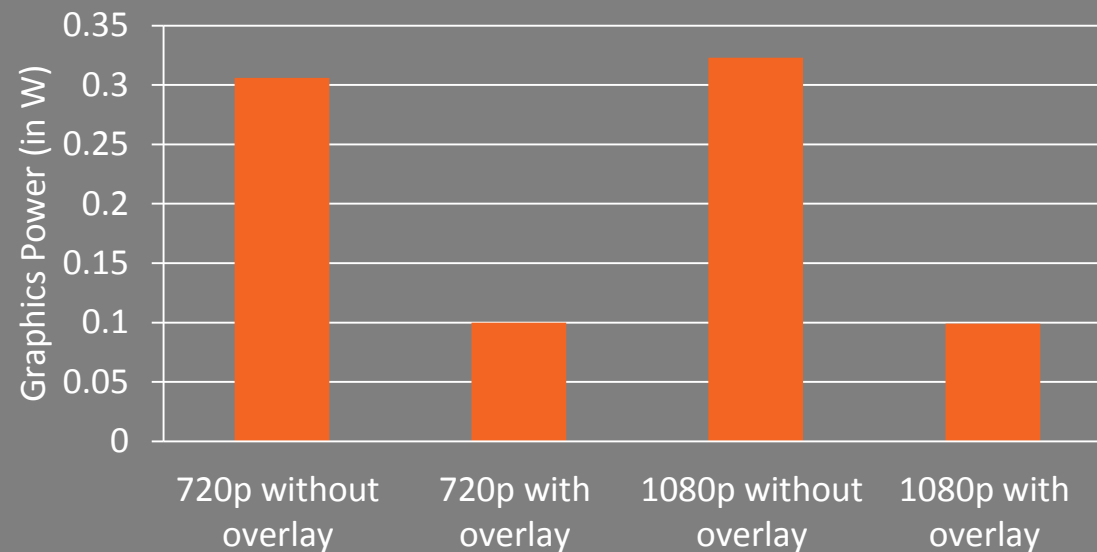


# OPTIMIZED VIDEO PLANE

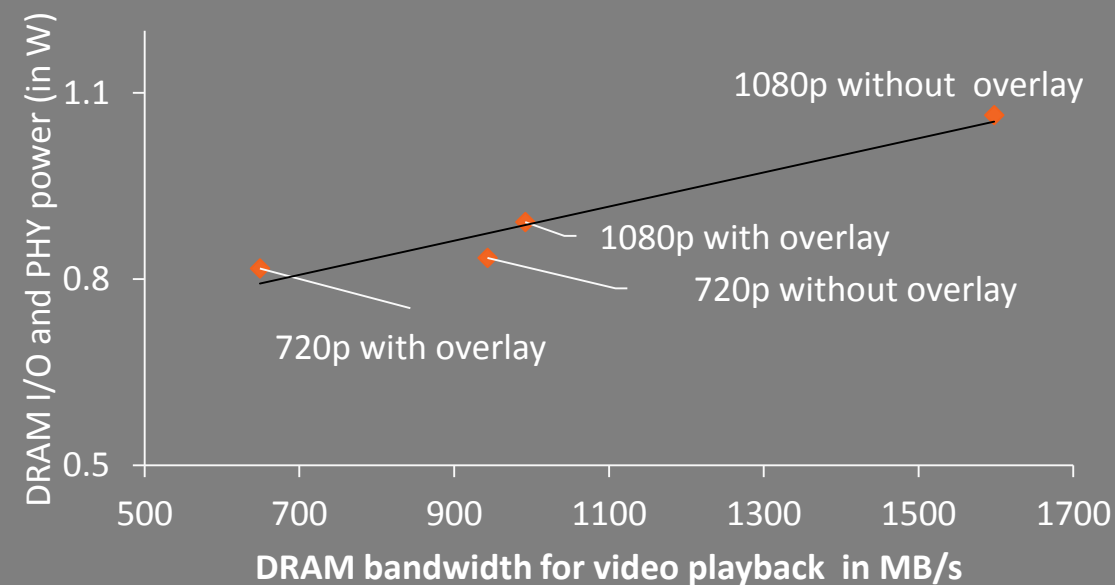
- ▲ Over 500mW of power savings with overlay feature or more than 20% of the overall use case power<sup>8</sup>
  - Over 200mW of power savings in graphics power plane by reducing the need for post-processing
  - Over 300mW of power savings in the DRAM subsystem by reducing bandwidth



Overlay power optimization in the Graphic voltage plane<sup>8</sup>



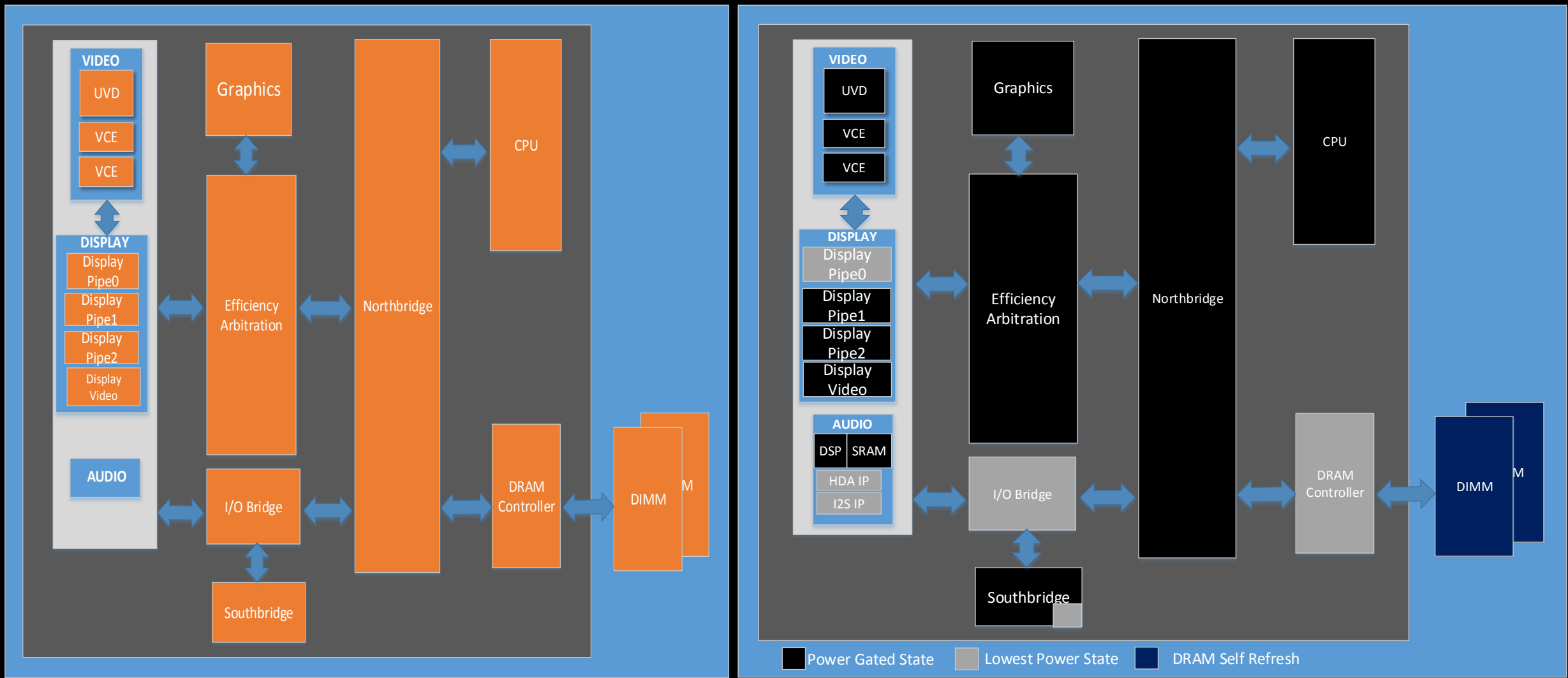
Overall power optimization in DRAM power<sup>8</sup>





# RACE TO DRAM SELF REFRESH

## TRANSITION FROM ACTIVE TO LOWEST POWER STATE



Active

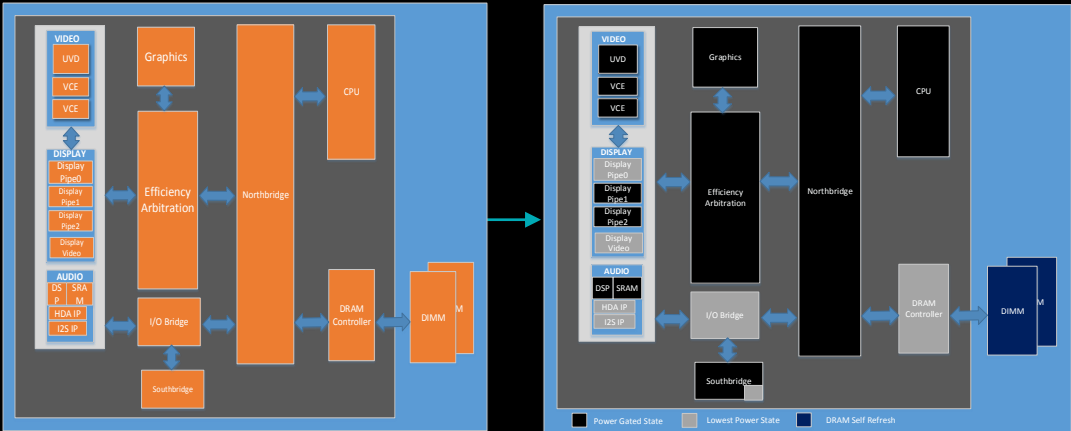
Lowest power state – DRAM in self-refresh

# INCREASED RESIDENCY IN DRAM SELF REFRESH

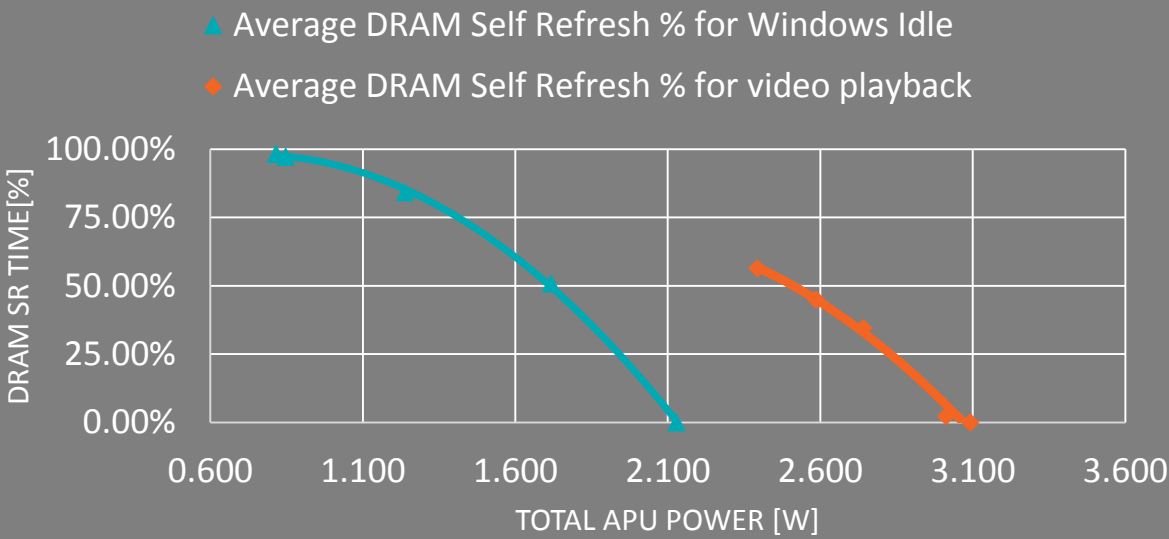


## RACE TO SELF-REFRESH

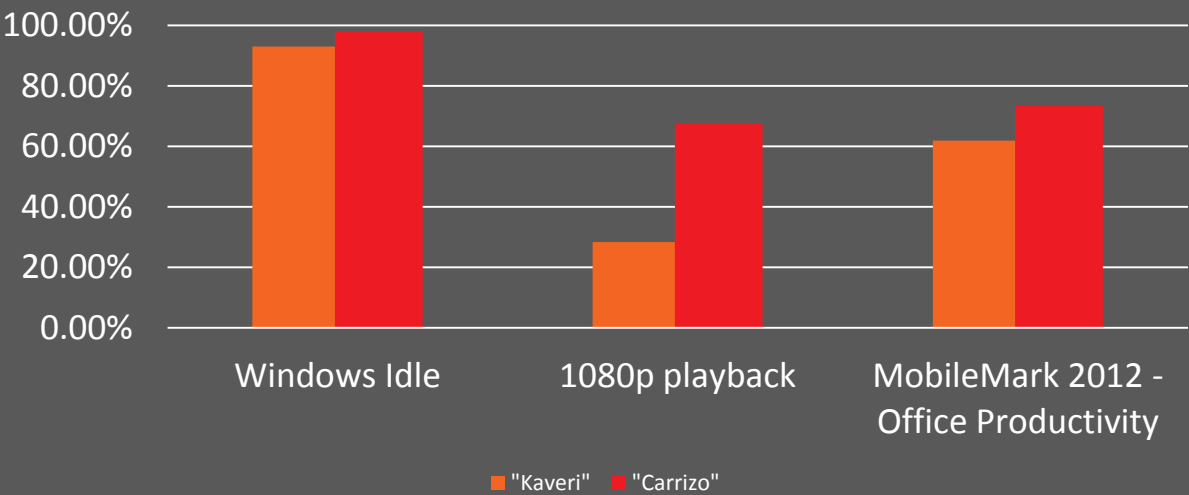
- ▲ New features in Carrizo puts SoC in the lowest power state and DRAM in self-refresh quicker than it did for the previous generation
  - GFX bandwidth compression
  - Decreased decode time in video playback
  - Reduced DRAM bandwidth with overlay plane
  - Integrated Southbridge



DRAM SELF REFRESH VS TOTAL APU POWER<sup>8</sup>



Average DRAM Self Refresh residency for Workloads<sup>18</sup>





*Enabling today.  
Inspiring tomorrow.*

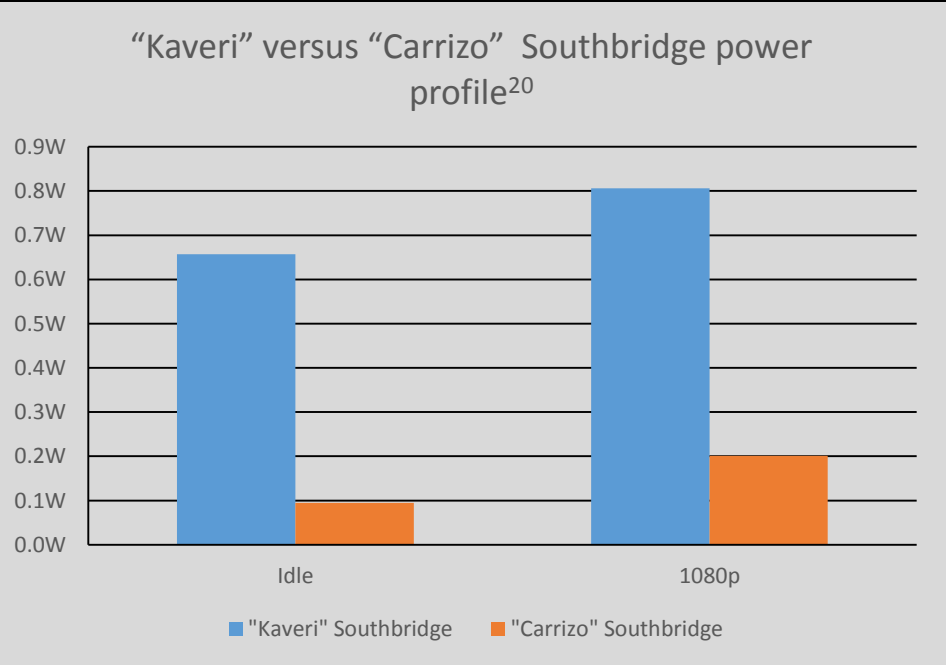
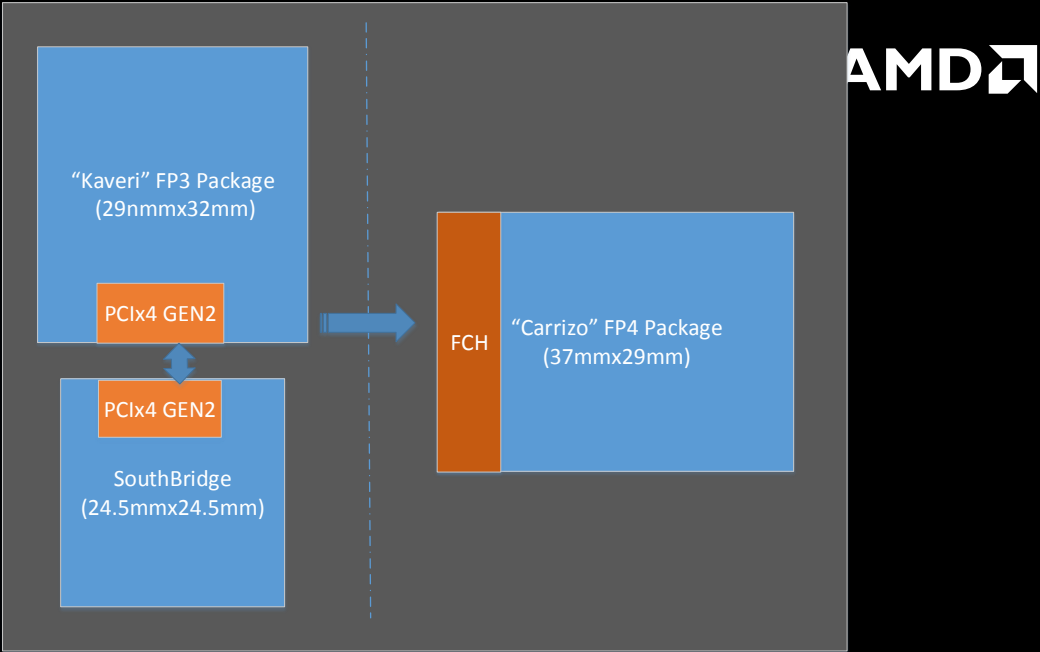
# OTHER POWER OPTIMIZATION

# ENERGY EFFICIENT SOUTHBRIDGE

## INTEGRATION BRINGS BATTERY BENEFITS

- ▲ Elimination of x4 GEN2 PCIE® link to discrete Southbridge chip set
  - Analog and controller power eliminated
  - Quick wake/shut down of the entire system
- ▲ Reduced voltage level for the main Southbridge core, analog IP
  - Core moved away from I/O voltage domain to a common voltage plane shared with other SoC IP
- ▲ Power gating of the Southbridge core IP

	6 <sup>th</sup> Generation APU "Carrizo" *	Previous Generation APU "Kaveri"*
Southbridge S0 IP	Variable (0.775V-1.15V)	1.1V
Southbridge Power gating	YES	NO
Southbridge ACPI S5 IP	0.775V	1.1V
Southbridge Analog IP	1.05V, 1.8V	1.1V, 3.3V



\* Nominal voltages. Details in AMD platform infrastructure specification

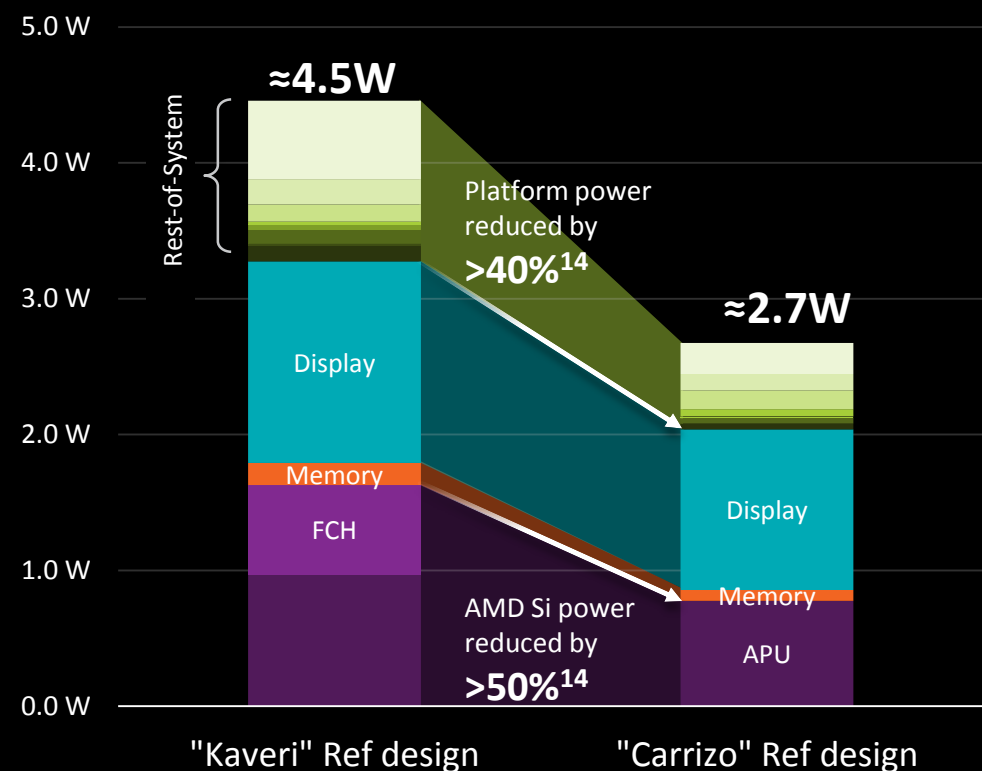


# SUMMARY: DRAMATIC BATTERY LIFE GAINS

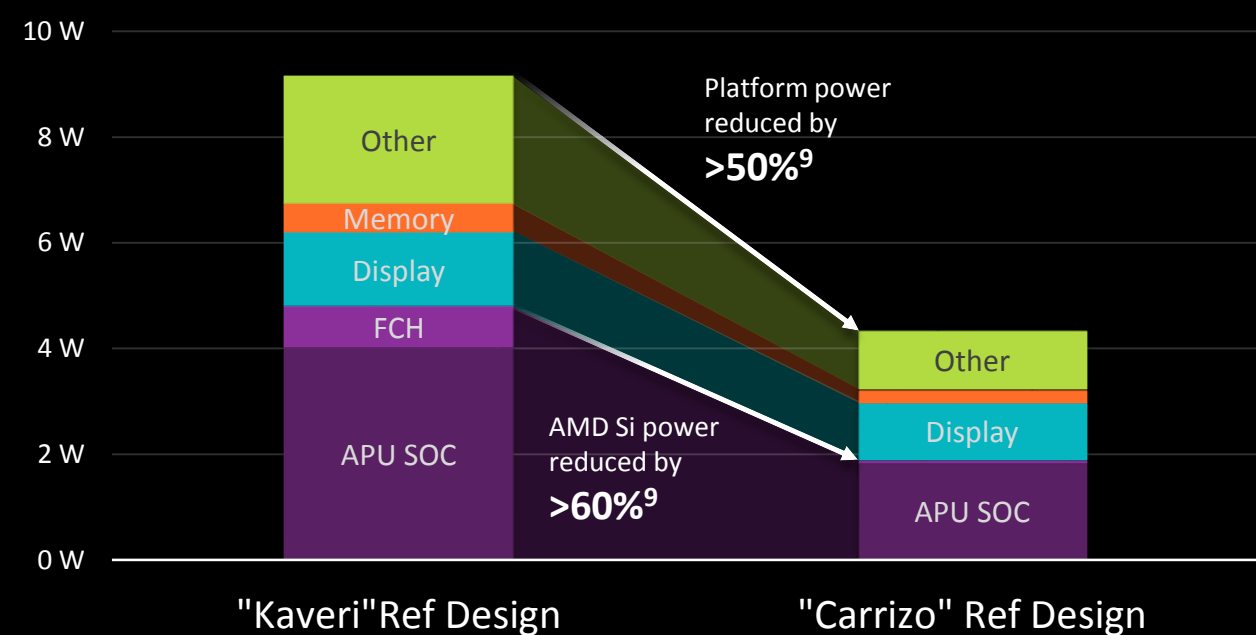


- ▲ 40%<sup>14</sup> platform power reduction for Windows Idle compared to the previous generation A-series APU
- ▲ Greater than 50% platform power reduction for HD video means a full doubling of battery life<sup>9</sup>
- ▲ 9.5 hours of HD video playback<sup>10</sup> means worry-free movie viewing on the go

Short Idle System Power



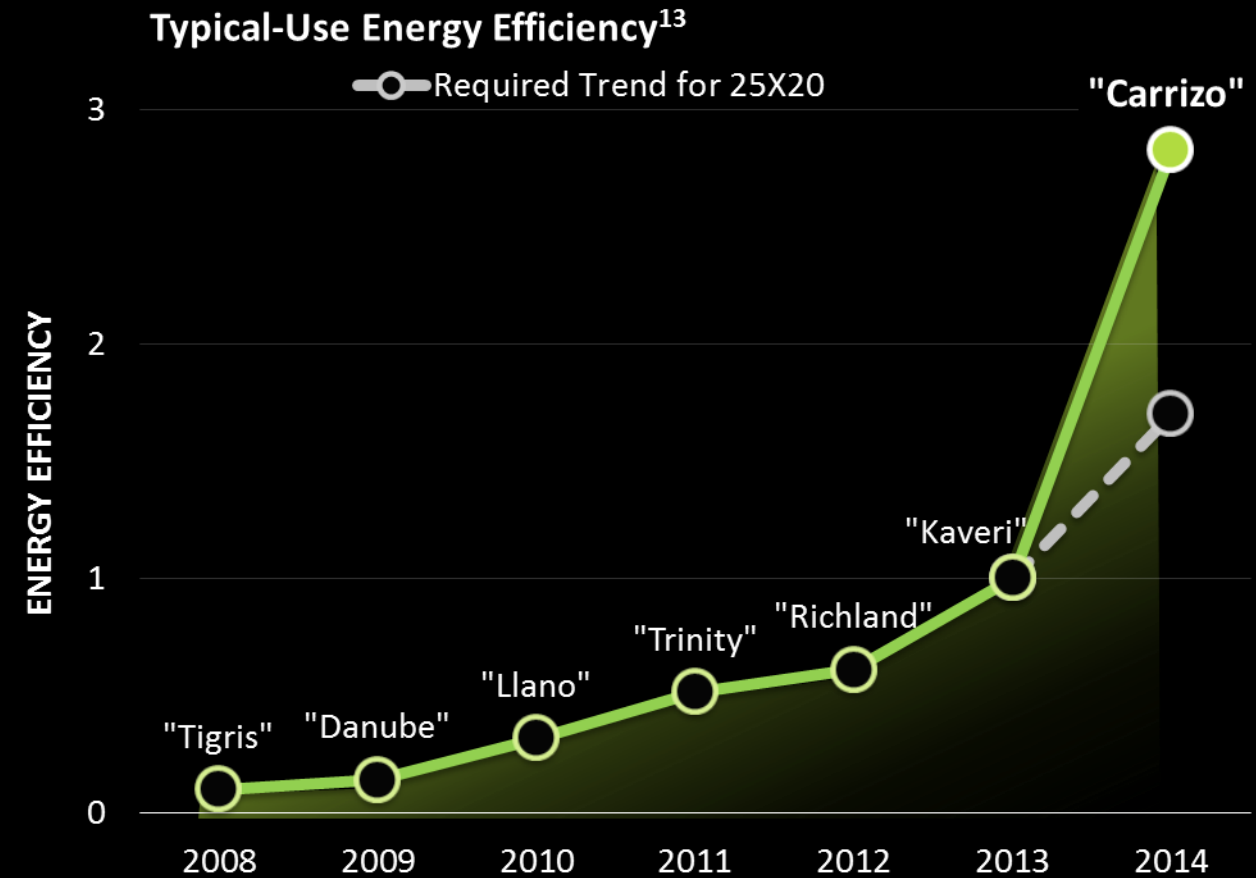
1080p Video Playback Power





# SUMMARY: IMPROVEMENTS IN ENERGY EFFICIENCY

- ▲ Innovated around architecture and implementation to enable compelling performance gains while remaining in a mature 28nm process node
  - 65% increase in 3DMark<sup>®</sup> 11
  - 350% reduction in decode time
  - 350% increase in transcoding throughput
  - H.265 & HEVC offload
- ▲ “Carrizo” provides a huge step towards AMD 25x20 initiative by delivering much better performance for much lower typical energy use



Typical power<sup>14</sup> reduced by  $\approx 2X$  while performance<sup>15</sup> increases up to almost  $1.5X^{16}$  = performance/Watt by  $2.4X^{17}$

1. A 28nm x86 APU Optimized for Power and Area Efficiency , presented by Kathryn Wilcox. Session 4.8, *Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* 2015 ISSCC Conference, February, 2015
2. AMD Lab measurements on AMD “Carrizo” GardeniaDAP FX-8800P (15W A1 DVT). 3D Mark Vantage performance measured with delta color compression enabled and disabled across 24 1920 x 1080 games with measured improvements between 5 and 7 percent.
3. AMD Lab measurements on AMD “Carrizo” GardeniaDAP system with 15W & 35W B10 (A1 DVT) at 36C Tambient with 3DMark®11 1.0.5 benchmark
4. AMD lab measurements on AMD “Carrizo” GardeniaDAP FX-8800P (15W (STAPM on) , DDR 1600 8GB and 35W A1 DVT , DDR 2133 8GB ) Win 8.1, 3DMark11-P 1.0.5 overall score versus “Kaveri” 19W and 35W reference designs, FX-7500 DDR 1600 8GB and DDR 2133 8GB respectively . “Carrizo” scored 1581 with STAPM on and 2120 with STAPM off, “Kaveri” scored 1220 with STAPM on and off on 15W platform. “Carrizo” scored 2500 and “Kaveri” scored 2184 on 35W platform.
5. AMD lab measurement on engineering samples with low bit rate HEVC video on Win8.1 using PowerDVD player to compare the total power differential between a CPU based decode and HW based decode.
6. Avatar 1080p clip running on fixed decode clock of 300Mhz, CPU clock of 2.5Ghz, NorthBridge clock of 1Ghz for 15W “Carrizo” and “Kaveri” with 512MB of frame buffer. Driver version Catalyst 14.12 & BIOS of WBL4709N for KV; 15.100.0 Beta 34 & BIOS of WGA5520N for CZ
7. AMD internal tool to measure transcode performance, in “speed” mode, with BBC, Space and Yozakura clips on engineering samples
8. AMD internal lab measurements on AMD “Carrizo” GardeniaDAP FX-8800P (15W A1 DVT) Win 8.1 1080p Big Buck Bunny on 720p and 1080p displays, BIOS WGA 5311M, Driver 15.10 Beta 30
9. AMD lab measurements on “Carrizo” 15W FX-8800P, 2x2GB DDR3L 1Rx16 SO-DIMMs, 14” 1366x760 Samsung LTN140AT29 display, 100 nits, 2.5” MM550 SSD, varibright enabled. “Kaveri” 19W reference design FX-7500, 2x4GB SO-DIMM, 14” 1366x768 100nit display, vari-bright enabled, 2.5” SATA SSD
10. AMD internal lab measurements on AMD “Carrizo” GardeniaDAP FX-8800P (15W A1 DVT) Win 8.1. PowerDVD, 1080p Big Buck Bunny video with 50WHr battery.
11. AMD lab measurements on AMD “Carrizo” GardeniaDAP FX-8800P as defined in the “Power Performance Operating Guide (PPOG)” v1.05 found in NDA.AMD.COM
12. AMD lab measurement on AMD “Carrizo” GardeniaDAP FX-8800P (15W A1 DVT) Win 8.1 1080p Big Buck Bunny on Windows Media Player with WGA5311N BIOS and 15.10.1020\_BR181142 driver; KV 15W (cTDP) with WBL4129N BIOS and 13.351 Beta 4 Build 0221 driver
13. Typical-use Energy Efficiency as defined by taking the ratio of compute capability as measured by common performance measures such as SpecIntRate, PassMark and PCMark, divided by typical energy use as defined by ETEC (Typical Energy Consumption for notebook computers) as specified in Energy Star Program Requirements Rev 6.0 10/2013.

14. AMD internal testing. Lab measurements on “Carrizo” 15WFX-8800P, 2x2GB DDR3L 1Rx16 SO-DIMMs, 14” 1366x768 Samsung LTN140AT29 display, 100 nits, 2.5” MM550 SSD, vari-bright enabled, 2.5” SATA SSD in Windows short-idle state “Kaveri” 19W reference design FX-7500, 2x4GB DDR3L 1Rx8 SO-DIMMs, 14” 1366x768 CMO, vari-bright enabled, 2.5” SATA SSD, Windows short-idle state.
15. AMD internal testing. “Carrizo” 15WFX-8800P, 2x2GB DDR3L 1Rx16 SO-DIMMs, 14” 1366x768 Samsung LTN140AT29 display, 100 nits, 2.5” MM550 SSD, vari-bright enabled, 2.5” SATA SSD. “Kaveri” 19W reference design FX-7500, 2x4GB DDR3L 1Rx8 SO-DIMMs, 14” 1366x768 CMO 100nit display, vari-bright enabled, 2.5” SATA SSD. Cinebench single-thread and multi-thread test results for frequency, instructions-per-clock and benchmarked performance
16. AMD lab measurements on 3.3 GHz Lab measurements on “Carrizo” 15WFX-8800P, 2x2GB DDR3L 1Rx16 SO-DIMMs, 14” 1366x768 Samsung LTN140AT29 display, 100 nits, 2.5” MM550 SSD, vari-bright enabled, 2.5” SATA SSD. “Kaveri” 19W reference design FX-7500, 2x4GB DDR3L 1Rx8 SO-DIMMs, 14” 1366x768 CMO, vari-bright enabled, 2.5” SATA SSD. Cinebench single-thread and multi-thread test results for frequency, instructions-per-clock and benchmarked performance
17. Typical-use Energy Efficiency as defined by taking the ratio of compute capability as measured by common performance measures such as SpecIntRate, PassMark and PCMark, divided by typical energy use as defined by ETEC (Typical Energy Consumption for notebook computers) as specified in Energy Star Program Requirements Rev 6.0 10/2013. “Kaveri” relative compute capability (4.5) of baseline divided by relative energy efficiency (0.45) of baseline = 10X. “Carrizo” relative compute capability (5.8) of baseline divided by relative energy efficiency (0.23) of baseline = 25.2X (which is 2.4x that of “Kaveri”)
18. Average DRAM self-refresh residency across two channels with configuration as defined by AMD Family 16h Models 30h-3Fh “Platform Performance and Power Optimization Guide” in [nda.amd.com](http://nda.amd.com) on a 15W CZ B10 FX 8800P TDP configuration and a 19W B10 FX-7500
19. AMD internal architecture model, 2013, to model typical Graphics bandwidth to memory system
20. AMD internal simulation testing 2013 for an internal Southbridge and compare versus lab measure power on external Southbridge on a KV reference board
21. AMD internal leakage model, 2013, for multi-media IP, with an assumption of 75% power gating of the “Carrizo” universal video decoder for average 1080p clip video playback

## DISCLAIMER

The information contained herein is for informational purposes only, and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

© 2015 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, and combinations thereof are trademarks of Advanced Micro Devices, Inc.

PCIe is a registered trademark of PCI-SIG Corporation.

3DMark is a trademark of Futuremark Corporation." and "MobileMark is a trademark of Business Applications Performance Corporation

Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies